

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:59:11 ; Search time 12.3077 Seconds
(without alignments)
83.892 Million cell updates/sec

Title: US-10-044-703-47

Perfect score: 96

Sequence: 1 NFLLPDAQSIQAAAAGFASK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:*
- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	351	4	US-09-118-426-6
2	96	100.0	373	4	US-09-118-426-5
3	96	100.0	374	3	US-08-818-112-153
4	96	100.0	374	4	US-08-818-111-148
5	96	100.0	374	4	US-09-056-556-153
6	96	100.0	374	4	US-09-056-556-155
7	96	100.0	374	4	US-09-072-596-148
8	96	100.0	374	4	US-09-072-596-150
9	96	100.0	374	4	US-09-072-596-153
10	96	100.0	374	4	US-09-072-596-155
11	96	100.0	374	4	US-09-287-849-6
12	96	100.0	374	4	US-09-287-849-40
13	96	100.0	652	4	US-09-072-596-350
14	96	100.0	652	4	US-09-072-596-355
15	96	100.0	802	4	US-09-056-556-214
16	96	100.0	802	4	US-09-072-596-209
17	96	100.0	802	4	US-09-072-596-346
18	96	100.0	802	4	US-09-072-596-214
19	96	100.0	802	4	US-09-072-596-351
20	96	100.0	802	4	US-09-287-849-10
21	45	46.9	243	4	US-09-543-681A-6371
22	44	45.8	490	4	US-09-489-039A-12361
23	44	45.8	7257	3	US-09-335-409-5
24	44	45.8	7257	4	US-09-568-102-5
25	44	45.8	7257	4	US-09-567-969-5
26	44	45.8	7257	4	US-09-568-480-5
27	44	45.8	7257	4	US-09-568-486-5

28	44	45.8	7257	4	US-09-568-472-5	Sequence 5, Appli
29	44	45.8	7257	4	US-09-567-899-5	Sequence 5, Appli
30	42	43.8	335	4	US-09-489-039A-8481	Sequence 8481, Ap
31	42	43.8	388	4	US-09-134-001C-3820	Sequence 3820, Ap
32	42	43.8	634	4	US-09-328-352-4627	Sequence 4627, Ap
33	41	42.7	259	4	US-09-252-991A-31123	Sequence 31123, A
34	41	42.7	380	4	US-09-252-991A-32854	Sequence 32854, A
35	41	42.7	730	3	US-08-872-757-2	Sequence 2, Appli
36	41	42.7	730	4	US-09-850-048A-2	Sequence 1, Appli
37	41	42.7	788	1	US-08-572-225-1	Sequence 1, Appli
38	41	42.7	986	4	US-09-285-385C-19	Sequence 19, Appli
39	40	41.7	358	4	US-09-252-991A-17017	Sequence 17017, A
40	40	41.7	378	1	US-07-723-002C-8	Sequence 8, Appli
41	40	41.7	494	4	US-09-489-039A-8154	Sequence 8154, Ap
42	40	41.7	661	4	US-09-540-236-3743	Sequence 3743, Ap
43	39.5	41.1	579	3	US-08-704-711A-1	Sequence 1, Appli
44	39.5	41.1	579	4	US-09-521-220-1	Sequence 1, Appli
45	39.5	41.1	582	3	US-08-704-711A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-118-426-6

; Sequence 6, Application US/09118426C

; Patent No. 6517839

; GENERAL INFORMATION:

; APPLICANT: Modlin, Robert L.

; APPLICANT: Libraty, Daniel H.

; TITLE OF INVENTION: METHODS FOR INDUCING INTERLEUKING-12 AND A TYPE 1/TH1

; TITLE OF INVENTION: T-CELL RESPONSE

; FILE REFERENCE: 30435.4US01

; CURRENT APPLICATION NUMBER: US/09/118,426C

; CURRENT FILING DATE: 1998-07-17

; EARLIER APPLICATION NUMBER: 60/052,970

; EARLIER FILING DATE: 1998-07-17

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 6

; LENGTH: 351

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; NAME/KEY: LIPID

; LOCATION: (1)

; OTHER INFORMATION: Xaa is N-ACYL DIGLYCERIDE cysteine

US-09-118-426-6

Query Match 100.0%; Score 96; DB 4; Length 351;

Best Local Similarity 100.0%; Pred.No. 7e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy *1 NFLLPDAQSIQAAAAGFASK 20

|||||

Db 249 NFLLPDAQSIQAAAAGFASK 269

RESULT 2

US-09-118-426-5

; Sequence 5, Application US/09118426C

; Patent No. 6517839

; GENERAL INFORMATION:

; APPLICANT: Modlin, Robert L.

; APPLICANT: Libraty, Daniel H.

; TITLE OF INVENTION: METHODS FOR INDUCING INTERLEUKING-12 AND A TYPE 1/TH1

; TITLE OF INVENTION: T-CELL RESPONSE

; FILE REFERENCE: 30435.4US01

; CURRENT APPLICATION NUMBER: US/09/118,426C

; CURRENT FILING DATE: 1998-07-17

; EARLIER APPLICATION NUMBER: 60/052,970

; EARLIER FILING DATE: 1998-07-17

; NUMBER OF SEQ ID NOS: 14

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; PUBLIC INFORMATION:
; JOURNAL: Infect. Immun.
; VOLUME: 57
; ISSUE: 8
; PAGES: 2481-
; DATE: 1989
; US-09-118-426-5

Query Match      100.0%; Score 96; DB 4; Length 373;
Best Local Similarity 100.0%; Pred. No. 7.5e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 NFLLPDAQSIQAAAAGFASK 20
Db      271 NFLLPDAQSIQAAAAGFASK 290

RESULT 3
US-08-818-112-153
; Sequence 153, Application US/08818112
; Patent No. 6290369
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 153:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-818-112-153

Query Match      100.0%; Score 96; DB 3; Length 374;
Best Local Similarity 100.0%; Pred. No. 7.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 NFLLPDAQSIQAAAAGFASK 20
Db      271 NFLLPDAQSIQAAAAGFASK 290

RESULT 4
US-08-818-111-148
; Sequence 148, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-818-111-148

Query Match      100.0%; Score 96; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 7.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 NFLLPDAQSIQAAAAGFASK 20
Db      272 NFLLPDAQSIQAAAAGFASK 291

RESULT 5
US-09-056-556-153
; Sequence 153, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; TREATME
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; US-09-056-556-153
```

```

;
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 153:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-056-556-153

Query Match 100.0%; Score 96; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 7.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFLLPDAQSIQAAAAGFASK 20
DB 272 NFLLPDAQSIQAAAAGFASK 291

RESULT 6
US-09-056-556-155
; Sequence 155, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 155:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; STRANDEDNESS:

```

```

; TOPOLOGY: linear
; US-09-056-556-155

Query Match 100.0%; Score 96; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 7.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFLLPDAQSIQAAAAGFASK 20
DB 272 NFLLPDAQSIQAAAAGFASK 291

RESULT 7
US-09-072-596-148
; Sequence 148, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-072-596-148

Query Match 100.0%; Score 96; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 7.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFLLPDAQSIQAAAAGFASK 20
DB 272 NFLLPDAQSIQAAAAGFASK 291

RESULT 8
US-09-072-596-150
; Sequence 150, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:

```

```

; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 150:
; LENGTH: 374 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-072-596-150

Query Match 100.0%; Score 96; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 7.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFFLPDAQSIQAAAAGFASK 20
Db 272 NFFLPDAQSIQAAAAGFASK 291

RESULT 9
US-09-072-967-153
; Sequence 153, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 150:
; LENGTH: 374 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-072-967-153

```

```

; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 153:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-072-967-153

Query Match 100.0%; Score 96; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 7.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFFLPDAQSIQAAAAGFASK 20
Db 272 NFFLPDAQSIQAAAAGFASK 291

RESULT 10
US-09-072-967-155
; Sequence 155, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 153:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-072-967-155

```


TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 155:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 374 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 US-09-072-967-155

Query Match 100.0%; Score 96; DB 4; Length 374;
 Best Local Similarity 100.0%; Pred. No. 7.6e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFLLPDAQSIQAAAGFASK 20
 Db 272 NFLLPDAQSIQAAAGFASK 291

RESULT 11

US-09-287-849-6
 Sequence 6, Application US/09287849
 Patent No. 6627198

GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, Davin C.

APPLICANT: Alderson, Mark

APPLICANT: Campos-Neto, Antonio
 APPLICANT: Corixa Corporation

TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 TITLE OF INVENTION: and Their Uses

FILE REFERENCE: 014058-009020US

CURRENT APPLICATION NUMBER: US/09/287,849
 CURRENT FILING DATE: 1999-04-07

PRIOR APPLICATION NUMBER: US 08/818,112

PRIOR FILING DATE: 1997-03-13

PRIOR APPLICATION NUMBER: US 08/942,578

PRIOR FILING DATE: 1997-10-01

PRIOR APPLICATION NUMBER: US 09/025,197

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 09/056,556

PRIOR FILING DATE: 1998-04-07

PRIOR APPLICATION NUMBER: US 09/223,040

PRIOR FILING DATE: 1998-12-30

NUMBER OF SEQ ID NOS: 46

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 6

LENGTH: 374

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: tri-fusion

US-09-287-849-6

Query Match 100.0%; Score 96; DB 4; Length 374;
 Best Local Similarity 100.0%; Pred. No. 7.6e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFLLPDAQSIQAAAGFASK 20
 Db 272 NFLLPDAQSIQAAAGFASK 291

RESULT 12

US-09-287-849-40

Sequence 40, Application US/09287849
 Patent No. 6627198

GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, Davin C.

APPLICANT: Alderson, Mark

APPLICANT: Campos-Neto, Antonio
 APPLICANT: Corixa Corporation
 TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 TITLE OF INVENTION: and Their Uses
 FILE REFERENCE: 014058-009020US
 CURRENT APPLICATION NUMBER: US/09/287,849
 CURRENT FILING DATE: 1999-04-07
 PRIOR APPLICATION NUMBER: US 08/818,112
 PRIOR FILING DATE: 1997-03-13
 PRIOR APPLICATION NUMBER: US 08/942,578
 PRIOR FILING DATE: 1997-10-01
 PRIOR APPLICATION NUMBER: US 09/025,197
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 09/056,556
 PRIOR FILING DATE: 1998-04-07
 PRIOR APPLICATION NUMBER: US 09/223,040
 PRIOR FILING DATE: 1998-12-30
 NUMBER OF SEQ ID NOS: 46
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 40
 LENGTH: 374
 TYPE: PRT
 ORGANISM: Mycobacterium tuberculosis
 FEATURE:
 OTHER INFORMATION: 38 kD antigen
 US-09-287-849-40

Query Match 100.0%; Score 96; DB 4; Length 374;
 Best Local Similarity 100.0%; Pred. No. 7.6e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFLLPDAQSIQAAAGFASK 20
 Db 272 NFLLPDAQSIQAAAGFASK 291

RESULT 13

US-09-072-596-350

Sequence 350, Application US/09072596
 Patent No. 6458366

GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, Davin C.

APPLICANT: Campos-Neto, Antonio
 APPLICANT: Houghton, Raymond

APPLICANT: Vedwick, Thomas S.
 APPLICANT: Twardzik, Daniel R.

APPLICANT: Lodes, Michael J.
 APPLICANT: Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 NUMBER OF SEQUENCES: 350

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/072,596

FILING DATE: 05-MAY-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.417C9

TELECOMMUNICATION INFORMATION:

TUBERCULOSIS

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:59:11 ; Search time 12.3077 Seconds
(without alignments)
83.892 Million cell updates/sec

Title: US-10-044-703-72

Perfect score: 109

Sequence: 1 YNINISLPSPYDPQKSLNY 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
- 2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*
- 3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*
- 4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*
- 5: /cgn2_6/prodata/2/iaa/6C COMB.pep.*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	228	3	US-08-569-221A-2
2	109	100.0	230	3	US-08-518-112-65
3	109	100.0	230	4	US-08-518-111-66
4	109	100.0	230	4	US-09-056-556-65
5	109	100.0	230	4	US-09-072-596-66
6	109	100.0	230	4	US-09-072-967-65
7	47	43.1	14	3	US-09-385-442-48
8	47	43.1	436	4	US-09-134-001C-3223
9	47	43.1	1264	1	US-07-789-915A-6
10	47	43.1	1264	1	US-08-005-002C-6
11	47	43.1	1264	1	US-08-487-203A-6
12	46	42.2	471	3	US-08-361-083-158
13	46	42.2	471	3	US-09-536-784-158
14	45	41.3	168	4	US-09-401-064-198
15	45	41.3	478	4	US-09-819-993-2
16	45	41.3	478	4	US-10-193-295-2
17	45	41.3	518	4	US-09-819-993-5
18	45	41.3	518	4	US-10-193-295-5
19	45	41.3	520	1	US-08-305-505-4
20	45	41.3	520	1	US-08-305-505-5
21	45	41.3	520	1	US-08-305-505-6
22	45	41.3	520	4	US-09-819-993-4
23	45	41.3	520	4	US-10-193-295-4
24	45	41.3	522	1	US-08-305-505-2
25	44	40.4	430	1	US-08-476-008-44
26	44	40.4	430	1	US-08-306-063-44
27	44	40.4	430	1	US-08-833-485-44

28	44	40.4	430	3	US-09-137-440-44	Sequence 44, Appl
29	43	39.4	209	4	US-09-489-039A-9030	Sequence 9030, Ap
30	43	39.4	237	4	US-09-489-039A-12839	Sequence 12839, A
31	43	39.4	669	4	US-09-361-631-7	Sequence 7, Appl
32	42	38.5	224	4	US-09-134-000C-5452	Sequence 5452, Ap
33	42	38.5	1041	3	US-08-898-978-2	Sequence 2, Appl
34	42	38.5	1041	4	US-09-372-858-2	Sequence 2, Appl
35	41.5	38.1	375	4	US-10-072-094-10	Sequence 10, Appl
36	41.5	38.1	441	4	US-09-134-000C-6703	Sequence 6703, Ap
37	41.5	38.1	557	4	US-09-134-000C-5061	Sequence 5061, Ap
38	41.5	38.1	562	4	US-09-134-000C-4308	Sequence 4308, Ap
39	41.5	38.1	571	4	US-09-071-035-168	Sequence 168, App
40	41.5	38.1	593	4	US-09-071-035-166	Sequence 166, App
41	41.5	38.1	757	4	US-09-585-858-23	Sequence 23, Appl
42	41.5	38.1	757	4	US-09-585-858-24	Sequence 24, Appl
43	41.5	38.1	1462	3	US-07-792-500-31	Sequence 31, Appl
44	41.5	38.1	1462	3	US-09-157-021-31	Sequence 31, Appl
45	41.5	38.1	1462	3	US-09-156-842-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-08-569-221A-2
; Sequence 2, Application US/08569221A
; Patent No. 6120776
; GENERAL INFORMATION:
; APPLICANT:
; NUMBER OF INVENTION: New diagnostic skin test for tuberculosis
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA: US/08/569, 221A
; FILING DATE: 28-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalski, Thomas J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 670001-2000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-569-221A-2

Query Match 100.0%; Score 109; DB 3; Length 228;
Best Local Similarity 100.0%; Pred. No. 5, 1e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YNINISLPSPYDPQKSLNY 20
Db 50 YNINISLPSPYDPQKSLNY 69

RESULT 2
US-08-818-112-65

; Sequence 65, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELEPHONE: (206) 682-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 230 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-65
; Query Match 100.0%; Score 109; DB 3; Length 230;
; Best Local Similarity 100.0%; Pred. No. 5.2e-10;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; Qy 1 YNINISLPSYYPDKSLNY 20
; Db 116 YNINISLPSYYPDKSLNY 135
; RESULT 3
; US-08-818-111-66
; Sequence 66, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA

; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELEPHONE: (206) 682-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 230 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-111-66
; Query Match 100.0%; Score 109; DB 4; Length 230;
; Best Local Similarity 100.0%; Pred. No. 5.2e-10;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; Qy 1 YNINISLPSYYPDKSLNY 20
; Db 116 YNINISLPSYYPDKSLNY 135
; RESULT 4
; US-09-056-556-65
; Sequence 65, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELEPHONE: (206) 682-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 230 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

TREATMI

US-09-056-556-65

Query Match 100.0%; Score 109; DB 4; Length 230;
Best Local Similarity 100.0%; Pred. No. 5.2e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YNINISLPSPYDPQKSLNY 20
Db 116 YNINISLPSPYDPQKSLNY 135

RESULT 5

US-09-072-596-66
; Sequence 66, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 230 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-072-596-66

Query Match 100.0%; Score 109; DB 4; Length 230;
Best Local Similarity 100.0%; Pred. No. 5.2e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YNINISLPSPYDPQKSLNY 20
Db 116 YNINISLPSPYDPQKSLNY 135

RESULT 6

US-09-072-967-65
; Sequence 65, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 230 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-072-967-65

Query Match 100.0%; Score 109; DB 4; Length 230;
Best Local Similarity 100.0%; Pred. No. 5.2e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YNINISLPSPYDPQKSLNY 20
Db 116 YNINISLPSPYDPQKSLNY 135

RESULT 7

US-09-385-442-48
; Sequence 48, Application US/09385442
; Patent No. 6200954
; GENERAL INFORMATION:
; APPLICANT: Ge, Ruowen
; APPLICANT: Kihl, R. Manjunatha
; TITLE OF INVENTION: Small Peptides Having Potent Anti-Angiogenic Activity
; FILE REFERENCE: 1781-170P
; CURRENT APPLICATION NUMBER: US/09/385,442
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 60/099,313
; EARLIER FILING DATE: 1999-09-04
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 48
; LENGTH: 14
; TYPE: PRT
; ORGANISM: mammalian
; FEATURE:
; OTHER INFORMATION: h/mFLK2

US-09-385-442-48

```
Query Match 43.1%; Score 47; DB 3; Length 14;
Best Local Similarity 58.3%; Pred. No. 0.21;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 NINISLPSYYPD 13
DB 3 NINVSPLARYPE 14

RESULT 8
US-09-134-001C-3223
; Sequence 3223, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3223
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3223

Query Match 43.1%; Score 47; DB 4; Length 436;
Best Local Similarity 47.1%; Pred. No. 14;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 INISLPSYYPQKSLEN 19
DB 417 VNVSPGFLPKLLEN 433

RESULT 9
US-07-789-915A-6
; Sequence 6, Application US/07789915A
; Patent No. 5212058
; GENERAL INFORMATION:
; APPLICANT: Baker, Rohan T.
; APPLICANT: Tobias, John W.
; APPLICANT: Varshavsky, Alexander
; TITLE OF INVENTION: Ubiquitin-Specific Proteases
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/789,915A
; FILING DATE: 19911108
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Broock, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: MIT-5091AA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
```

```
TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1264 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-789-915A-6

Query Match 43.1%; Score 47; DB 1; Length 1264;
Best Local Similarity 44.4%; Pred. No. 54;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 INISLPSYYPQKSLENY 20
DB 449 INLSVSHYTDRIIRNY 466

RESULT 10
US-08-005-002C-6
; Sequence 6, Application US/08005002C
; Patent No. 5434818
; GENERAL INFORMATION:
; APPLICANT: Baker, Rohan T.
; APPLICANT: Tobias, John W.
; APPLICANT: Varshavsky, Alexander
; TITLE OF INVENTION: Ubiquitin-Specific Proteases
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: Maine
; COUNTRY: U.S.A.
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/005,002C
; FILING DATE: 15-JAN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/789,915
; FILING DATE: 08-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: MIT-5091AAZ
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 207-363-0558
; TELEFAX: 207-363-0528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1264 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-005-002C-6

Query Match 43.1%; Score 47; DB 1; Length 1264;
Best Local Similarity 44.4%; Pred. No. 54;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 INISLPSYYPQKSLENY 20
DB 449 INLSVSHYTDRIIRNY 466

RESULT 11
US-08-487-203A-6
```

```
; Sequence 6, Application US/08487203A
; Patent No. 5683904
; GENERAL INFORMATION:
; APPLICANT: Baker, Rohan T.
; APPLICANT: Tobias, John W.
; APPLICANT: Varshavsky, Alexander
; TITLE OF INVENTION: Ubiquitin-Specific Proteases
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: Maine
; COUNTRY: U.S.A.
; ZIP: 03911
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,203A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/005,002
; FILING DATE: 15-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: MIT-5091A3Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 207-363-0558
; TELEFAX: 207-363-0528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1264 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-487-203A-6

Query Match 43.1%; Score 47; DB 1; Length 1264;
Best Local Similarity 44.4%; Pred. No. 54;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 3 INISLPSYYPDQKSLNY 20
|||: |||: |||
Db 449 INLSVSHYVTRDRI 466

RESULT 12
US-08-961-083-158
; Sequence 158, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE: OCT-30-1997
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 158:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 158:
US-09-536-784-158
```

```
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 158:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-158

Query Match 42.2%; Score 46; DB 3; Length 471;
Best Local Similarity 45.0%; Pred. No. 23;
Matches 9; Conservative 4; Mismatches 3; Indels 4; Gaps 1;

Qy 1 YNINISLPSYYPDQKSLNY 20
|||: |||: |||
Db 355 YNNWLSYPIYY----DVENW 370

RESULT 13
US-09-536-784-158
; Sequence 158, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 158:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 471 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 158:
US-09-536-784-158
```

Search completed: March 10, 2004, 12:14:17
Job time : 13:3077 secs

Query Match 42.2%; Score 46; DB 4; Length 471;
Best Local Similarity 45.0%; Pred. No. 23;
Matches 9; Conservative 4; Mismatches 3; Indels 4; Gaps 1;

QY 1 YNINISLPSYYPDQKSLNY 20
Db 355 YNNLSYPIYI---DVENW 370

RESULT 14
US-09-401-064-198
; Sequence 198; Application US/09401064
; Patent No. 6623923
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeline Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tonglong
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C2
; CURRENT APPLICATION NUMBER: US/09/401,064
; CURRENT FILING DATE: 1999-09-22
; NUMBER OF SEQ ID NOS: 371
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 198
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-401-064-198

Query Match 41.3%; Score 45; DB 4; Length 168;
Best Local Similarity 50.0%; Pred. No. 9.4;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 3 INISLPSYYPDQKSLNY 20
Db 60 LEIYFSPQYVDQAELEKY 77

RESULT 15
US-09-819-993-2
; Sequence 2; Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Human
US-09-819-993-2

Query Match 41.3%; Score 45; DB 4; Length 478;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 3 INISLPSYYPDQKSLNY 20
Db 22 LEIYFSPQYVDQAELEKY 39

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:59:11 ; Search time 10.4615 Seconds
(without alignments)
83.892 Million cell updates/sec

Title: US-10-044-703-49

Perfect score: 81

Sequence: 1 TGSGAGIAQAAGTVNI 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	100.0	351	4	US-09-118-426-6
2	81	100.0	373	4	US-09-118-426-5
3	81	100.0	374	3	US-08-818-112-153
4	81	100.0	374	4	US-08-818-111-148
5	81	100.0	374	4	US-09-056-556-133
6	81	100.0	374	4	US-09-056-556-135
7	81	100.0	374	4	US-09-072-967-148
8	81	100.0	374	4	US-09-072-967-150
9	81	100.0	374	4	US-09-072-967-153
10	81	100.0	374	4	US-09-072-967-155
11	81	100.0	374	4	US-09-287-849-6
12	81	100.0	374	4	US-09-287-849-40
13	81	100.0	652	4	US-09-072-967-350
14	81	100.0	652	4	US-09-072-967-355
15	81	100.0	802	4	US-09-056-556-214
16	81	100.0	802	4	US-09-072-967-209
17	81	100.0	802	4	US-09-072-967-346
18	81	100.0	802	4	US-09-072-967-351
19	81	100.0	802	4	US-09-072-967-354
20	81	100.0	802	4	US-09-287-849-10
21	47	58.0	301	4	US-09-107-532A-4953
22	46	56.8	246	4	US-09-134-000C-3407
23	45	55.6	261	4	US-09-071-035-140
24	45	55.6	284	4	US-09-071-035-138
25	45	55.6	289	4	US-09-107-532A-4402
26	45	55.6	306	4	US-09-134-000C-5264
27	45	55.6	1184	4	US-10-153-064-89

28	44	54.3	27	1	US-08-425-069-59	Sequence 59, Appl
29	44	54.3	27	2	US-08-317-848B-59	Sequence 59, Appl
30	44	54.3	231	4	US-09-252-991A-22801	Sequence 22801, A
31	44	54.3	651	3	US-08-556-978B-19	Sequence 19, Appl
32	44	54.3	651	3	US-09-247-806-1	Sequence 1, Appl
33	44	54.3	651	4	US-09-863-859-1	Sequence 2, Appl
34	44	54.3	718	1	US-08-425-069-2	Sequence 1, Appl
35	44	54.3	718	2	US-08-317-848B-2	Sequence 2, Appl
36	44	54.3	747	3	US-09-034-177-3	Sequence 3, Appl
37	43	53.1	251	4	US-09-252-991A-32676	Sequence 32676, A
38	43	53.1	469	4	US-09-252-991A-25438	Sequence 25438, A
39	43	53.1	538	4	US-09-252-991A-21622	Sequence 21622, A
40	43	53.1	832	1	US-08-209-747-2	Sequence 2, Appl
41	43	53.1	832	1	US-08-458-298-2	Sequence 2, Appl
42	42	51.9	27	1	US-08-425-069-51	Sequence 51, Appl
43	42	51.9	27	2	US-08-317-848B-51	Sequence 51, Appl
44	42	51.9	27	3	US-08-556-978B-29	Sequence 29, Appl
45	42	51.9	27	3	US-08-556-978B-92	Sequence 92, Appl

ALIGNMENTS

RESULT 1

US-09-118-426-6

; Sequence 6, Application US/09118426C

; Patent No. 6517839

; GENERAL INFORMATION:

; APPLICANT: Modlin, Robert L.

; APPLICANT: Libracy, Daniel H.

; TITLE OF INVENTION: METHODS FOR INDUCING INTERLEUKING-12 AND A TYPE 1/TH1

; FILE REFERENCE: 30435.4US01

; CURRENT APPLICATION NUMBER: US/09/118,426C

; CURRENT FILING DATE: 1998-07-17

; EARLIER APPLICATION NUMBER: 60/052,970

; EARLIER FILING DATE: 1998-07-17

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 6

; LENGTH: 351

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; NAME/KEY: LIPID

; LOCATION: (1)

; OTHER INFORMATION: Xaa is N-ACYL DIGLYCERIDE cysteine

US-09-118-426-6

Query Match 100.0%; Score 81; DB 4; Length 351;

Best Local Similarity 100.0%; Pred No. 0.00039; Indels 0; Gaps 0;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGSGAGIAQAAGTVNI 17

Db 63 TGSGAGIAQAAGTVNI 79

RESULT 2

US-09-118-426-5

; Sequence 5, Application US/09118426C

; Patent No. 6517839

; GENERAL INFORMATION:

; APPLICANT: Modlin, Robert L.

; APPLICANT: Libracy, Daniel H.

; TITLE OF INVENTION: METHODS FOR INDUCING INTERLEUKING-12 AND A TYPE 1/TH1

; FILE REFERENCE: 30435.4US01

; CURRENT APPLICATION NUMBER: US/09/118,426C

; CURRENT FILING DATE: 1998-07-17

; EARLIER APPLICATION NUMBER: 60/052,970

; EARLIER FILING DATE: 1998-07-17

; NUMBER OF SEQ ID NOS: 14

```
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 5
/ LENGTH: 373
/ TYPE: PRT
/ ORGANISM: Mycobacterium tuberculosis
/ PUBLICATION INFORMATION:
/ JOURNAL: Infect. Immun.
/ VOLUME: 57
/ ISSUE: 8
/ PAGES: 2481-
/ DATE: 1989
/ US-09-118-426-5

Query Match          100.0%; Score 81; DB 4; Length 373;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGSGAGIAQAAAGTVNI 17
Db 85 TGSGAGIAQAAAGTVNI 101

RESULT 3
US-08-818-112-153
/ Sequence 153, Application US/08818112
/ Patent No. 6290969
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, David C.
/ APPLICANT: Campos-Neto, Antonio
/ APPLICANT: Houghton, Raymond
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Twardzik, Daniel R.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
/ TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
/ NUMBER OF SEQUENCES: 153
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED and BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: USA
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/818,112
/ FILING DATE: 13-MAR-1997
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.411C6
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 582-6031
/ INFORMATION FOR SEQ ID NO: 153:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 374 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-818-112-153

Query Match          100.0%; Score 81; DB 3; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGSGAGIAQAAAGTVNI 17
Db 85 TGSGAGIAQAAAGTVNI 101

RESULT 4
US-08-818-111-148
/ Sequence 148, Application US/08818111
/ Patent No. 6338852
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, David C.
/ APPLICANT: Campos-Neto, Antonio
/ APPLICANT: Houghton, Raymond
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Twardzik, Daniel R.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
/ TITLE OF INVENTION: TUBERCULOSIS
/ NUMBER OF SEQUENCES: 148
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED and BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: USA
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/818,111
/ FILING DATE: 13-MAR-1997
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.417C6
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 582-6031
/ INFORMATION FOR SEQ ID NO: 148:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 374 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-818-111-148

Query Match          100.0%; Score 81; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGSGAGIAQAAAGTVNI 17
Db 86 TGSGAGIAQAAAGTVNI 102

RESULT 5
US-09-056-556-153
/ Sequence 153, Application US/09056556
/ Patent No. 6350456
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, David C.
/ APPLICANT: Houghton, Raymond
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Twardzik, Daniel R.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
/ TITLE OF INVENTION: TUBERCULOSIS
/ NUMBER OF SEQUENCES: 241
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED and BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ US-09-056-556-153

Query Match          100.0%; Score 81; DB 3; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGSGAGIAQAAAGTVNI 17
Db 85 TGSGAGIAQAAAGTVNI 101
```

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 153:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-056-556-153

Query Match 100.0%; Score 81; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSGAGIAQAAAGTVNI 17
DB 86 TGSGAGIAQAAAGTVNI 102

RESULT 6
US-09-056-556-155
Sequence 155, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 155:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
STRANDEDNESS:

TOPOLOGY: linear
US-09-056-556-155
Query Match 100.0%; Score 81; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSGAGIAQAAAGTVNI 17
DB 86 TGSGAGIAQAAAGTVNI 102

RESULT 7
US-09-072-596-148
Sequence 148, Application US/09072596
Patent No. 6458366
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedwick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 148:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-072-596-148

Query Match 100.0%; Score 81; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSGAGIAQAAAGTVNI 17
DB 86 TGSGAGIAQAAAGTVNI 102

RESULT 8
US-09-072-596-150
Sequence 150, Application US/09072596
Patent No. 6458366
GENERAL INFORMATION:

```

; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodges, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; INFORMATION FOR SEQ ID NO: 150:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; STRANDEDNESS:
; US-09-072-596-150

Query Match 100.0%; Score 81; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGSGAGIAQAAAGTVNI 17
Db 86 TGSGAGIAQAAAGTVNI 102

RESULT 9
US-09-072-967-153
; Sequence 153, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodges, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:

```

```

; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 153:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-072-967-153

Query Match 100.0%; Score 81; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGSGAGIAQAAAGTVNI 17
Db 86 TGSGAGIAQAAAGTVNI 102

RESULT 10
US-09-072-967-155
; Sequence 155, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodges, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 155:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-072-967-155

Query Match 100.0%; Score 81; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSGGAGIAQAAAGTVNI 17
   |||||
DB 86 TSGGAGIAQAAAGTVNI 102

RESULT 11
US-09-287-849-6
; Sequence 6, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 6
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-6

Query Match 100.0%; Score 81; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSGGAGIAQAAAGTVNI 17
   |||||
DB 86 TSGGAGIAQAAAGTVNI 102

RESULT 12
US-09-287-849-40
; Sequence 40, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark

```

```

; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 40
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: 38 kD antigen
US-09-287-849-40

Query Match 100.0%; Score 81; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSGGAGIAQAAAGTVNI 17
   |||||
DB 86 TSGGAGIAQAAAGTVNI 102

RESULT 13
US-09-072-596-350
; Sequence 350, Application US/09072596
; Patent No. 8458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 350:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-072-596-350

Query Match 100.0%; Score 81; DB 4; Length 652;
Best Local Similarity 100.0%; Pred. No. 0.00072; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0;

Qy 1 TSGGAGIAQAAGTVNI 17
Db 70 TSGGAGIAQAAGTVNI 86

RESULT 14

US-09-072-967-355
Sequence 355, Application US/09072967
Patent No. 6592877
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedwick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 355:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-072-967-355

Query Match 100.0%; Score 81; DB 4; Length 652;
Best Local Similarity 100.0%; Pred. No. 0.00072; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0;

Qy 1 TSGGAGIAQAAGTVNI 17
Db 70 TSGGAGIAQAAGTVNI 86

RESULT 15

US-09-056-556-214
Sequence 214, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 214:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-214

Query Match 100.0%; Score 81; DB 4; Length 802;
Best Local Similarity 100.0%; Pred. No. 0.00089; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0;

Qy 1 TSGGAGIAQAAGTVNI 17
Db 136 TSGGAGIAQAAGTVNI 152

Search completed: March 10, 2004, 12:14:14
Job time : 10.4615 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:59:11 ; Search time 12.3077 Seconds
(without alignments)
83.892 Million cell updates/sec

Title: US-10-044-703-52

Perfect score: 103

Sequence: 1 AETPGCVAYIGISFLDQASQ 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/2/aaa/5A COMB.pap:*

2: /cgn2_6/prodata/2/aaa/5B COMB.pap:*

3: /cgn2_6/prodata/2/aaa/6A COMB.pap:*

4: /cgn2_6/prodata/2/aaa/6B COMB.pap:*

5: /cgn2_6/prodata/2/aaa/PCTUS COMB.pap:*

6: /cgn2_6/prodata/2/aaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	103	100.0	351	4	US-09-118-426-6
2	103	100.0	373	4	US-09-118-426-5
3	103	100.0	374	3	US-08-818-112-153
4	103	100.0	374	4	US-08-818-111-148
5	103	100.0	374	4	US-09-056-556-153
6	103	100.0	374	4	US-09-056-556-155
7	103	100.0	374	4	US-09-072-596-148
8	103	100.0	374	4	US-09-072-596-150
9	103	100.0	374	4	US-09-072-967-153
10	103	100.0	374	4	US-09-072-967-155
11	103	100.0	374	4	US-09-287-849-6
12	103	100.0	374	4	US-09-287-849-40
13	103	100.0	652	4	US-09-072-596-350
14	103	100.0	652	4	US-09-072-967-355
15	103	100.0	802	4	US-09-056-556-214
16	103	100.0	802	4	US-09-072-596-209
17	103	100.0	802	4	US-09-072-596-346
18	103	100.0	802	4	US-09-072-967-214
19	103	100.0	802	4	US-09-072-967-351
20	103	100.0	802	4	US-09-287-849-10
21	58	56.3	301	4	US-09-107-532A-4953
22	50	48.5	246	4	US-09-134-000C-3407
23	46	44.7	289	4	US-09-107-532A-4402
24	46	44.7	480	4	US-09-489-039A-9157
25	44	42.7	124	4	US-08-858-207A-380
26	44	42.7	175	4	US-09-252-991A-25228
27	44	42.7	152	4	US-09-198-452A-445

28 43 41.7 320 4 US-09-252-991A-20760 Sequence 20760, A
29 42 40.8 101 4 US-09-328-352-6449 Sequence 6449, Ap
30 42 40.8 389 4 US-09-107-532A-6456 Sequence 6456, Ap
31 42 40.8 432 4 US-08-311-731A-251 Sequence 251, App
32 42 40.8 545 4 US-09-269-731-8 Sequence 8, Appli
33 42 40.8 1449 3 US-08-840-062-6 Sequence 6, Appli
34 42 40.8 1722 4 US-09-194-612A-1 Sequence 1, Appli
35 42 40.8 1723 4 US-09-194-612A-31 Sequence 31, Appli
36 41 39.8 288 4 US-09-198-452A-593 Sequence 593, App
37 41 39.8 456 4 US-09-134-000C-5403 Sequence 5403, Ap
38 40.5 39.3 135 2 US-08-647-960-3 Sequence 3, Appli
39 40 38.8 183 4 US-09-489-039A-11620 Sequence 11620, A
40 40 38.8 428 4 US-09-198-452A-720 Sequence 720, App
41 40 38.8 574 4 US-09-079-431B-6 Sequence 6, Appli
42 40 38.8 625 4 US-08-759-436-3 Sequence 3, Appli
43 40 38.8 625 4 US-08-759-436-5 Sequence 5, Appli
44 40 38.8 629 4 US-09-079-431B-4 Sequence 4, Appli
45 40 38.8 630 4 US-09-079-431B-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-118-426-6

; Sequence 6, Application US/09118426C

; Patent No. 6517839

; GENERAL INFORMATION:

; APPLICANT: Modlin, Robert L.

; APPLICANT: Libraty, Daniel H.

; TITLE OF INVENTION: METHODS FOR INDUCING INTERLEUKING-12 AND A TYPE 1/TH1

; TITLE OF INVENTION: T-CELL RESPONSE

; FILE REFERENCE: 30435.4US01

; CURRENT APPLICATION NUMBER: US/09/118,426C

; CURRENT FILING DATE: 1998-07-17

; EARLIER APPLICATION NUMBER: 60/052,970

; EARLIER FILING DATE: 1998-07-17

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 6

; LENGTH: 351

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; NAME/KEY: LIPID

; LOCATION: (1)

; OTHER INFORMATION: Xaa is N-ACYL DIGLYCERIDE cysteine

US-09-118-426-6

Query Match 100.0%; Score 103; DB 4; Length 351;

Best Local Similarity 100.0%; Pred. No. 1e-09; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AETPGCVAYIGISFLDQASQ 20

|||||

Db 216 AETPGCVAYIGISFLDQASQ 235

RESULT 2

US-09-118-426-5

; Sequence 5, Application US/09118426C

; Patent No. 6517839

; GENERAL INFORMATION:

; APPLICANT: Modlin, Robert L.

; APPLICANT: Libraty, Daniel H.

; TITLE OF INVENTION: METHODS FOR INDUCING INTERLEUKING-12 AND A TYPE 1/TH1

; TITLE OF INVENTION: T-CELL RESPONSE

; FILE REFERENCE: 30435.4US01

; CURRENT APPLICATION NUMBER: US/09/118,426C

; CURRENT FILING DATE: 1998-07-17

; EARLIER APPLICATION NUMBER: 60/052,970

; EARLIER FILING DATE: 1998-07-17

; NUMBER OF SEQ ID NOS: 14

```
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 5
/ LENGTH: 373
/ TYPE: PRT
/ ORGANISM: Mycobacterium tuberculosis
/ PUBLICATION INFORMATION:
/ JOURNAL: Infect. Immun.
/ VOLUME: 57
/ ISSUE: 8
/ PAGES: 2481-
/ DATE: 1989
/
US-09-118-426-5

Query Match 100.0%; Score 103; DB 4; Length 373;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AETPGCVAYIGISFLDQASQ 20
Db 238 AETPGCVAYIGISFLDQASQ 257

RESULT 3
US-08-818-112-153
/ Sequence 153, Application US/08818112
/ Patent No. 6290969
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Campos-Neto, Antonio
/ APPLICANT: Houghton, Raymond
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Twardzik, Daniel R.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
/ TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
/ NUMBER OF SEQUENCES: 153
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED and BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: USA
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ FILING DATE: 13-MAR-1997
/ APPLICATION NUMBER: US/08/818,112
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.411C6
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 153:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 374 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
US-08-818-112-153

Query Match 100.0%; Score 103; DB 3; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AETPGCVAYIGISFLDQASQ 20
Db 238 AETPGCVAYIGISFLDQASQ 257

RESULT 4
US-08-818-111-148
/ Sequence 148, Application US/08818111
/ Patent No. 6338852
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Campos-Neto, Antonio
/ APPLICANT: Houghton, Raymond
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Twardzik, Daniel R.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
/ TITLE OF INVENTION: TUBERCULOSIS
/ NUMBER OF SEQUENCES: 148
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED and BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: USA
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ FILING DATE: 13-MAR-1997
/ APPLICATION NUMBER: US/08/818,111
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.417C6
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 148:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 374 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
US-08-818-111-148

Query Match 100.0%; Score 103; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AETPGCVAYIGISFLDQASQ 20
Db 239 AETPGCVAYIGISFLDQASQ 258

RESULT 5
US-09-056-556-153
/ Sequence 153, Application US/09056556
/ Patent No. 6350456
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, Davin C.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
/ TITLE OF INVENTION: TUBERCULOSIS
/ NUMBER OF SEQUENCES: 241
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED and BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/
US-09-056-556-153

Query Match 100.0%; Score 103; DB 3; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AETPGCVAYIGISFLDQASQ 20
Db 239 AETPGCVAYIGISFLDQASQ 258

RESULT 6
US-08-818-112-153
/ Sequence 153, Application US/08818112
/ Patent No. 6290969
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Campos-Neto, Antonio
/ APPLICANT: Houghton, Raymond
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Twardzik, Daniel R.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
/ TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
/ NUMBER OF SEQUENCES: 153
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED and BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ FILING DATE: 13-MAR-1997
/ APPLICATION NUMBER: US/08/818,112
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.411C6
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 153:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 374 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
US-08-818-112-153
```

TREATME


```

; TOPOLOGY: linear
; US-09-056-556-155

Query Match 100.0%; Score 103; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAVIGISFLDQASQ 20
| | | | | | | | | | | | | | | | | | | | | |
Db 239 AETPGCVAVIGISFLDQASQ 258

RESULT 7
US-09-072-596-148
; Sequence 148, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 522-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-072-596-148

Query Match 100.0%; Score 103; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAVIGISFLDQASQ 20
| | | | | | | | | | | | | | | | | | | | | |
Db 239 AETPGCVAVIGISFLDQASQ 258

RESULT 8
US-09-072-596-150
; Sequence 150, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:

```

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 150:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-150
Query Match 100.0%; Score 103; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AETPGCVAYIGISFLDQASQ 20
Db 239 AETPGCVAYIGISFLDQASQ 258
RESULT 9
US-09-072-967-153
Sequence 153, Application US/09072967
Patent No. 6592877
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 153:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-072-967-153
Query Match 100.0%; Score 103; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AETPGCVAYIGISFLDQASQ 20
Db 239 AETPGCVAYIGISFLDQASQ 258
RESULT 10
US-09-072-967-155
Sequence 155, Application US/09072967
Patent No. 6592877
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 350:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-072-967-350

Query Match 100.0%; Score 103; DB 4; Length 652;
Best Local Similarity 100.0%; Pred. No. 2e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
DB 223 AETPGCVAYIGISFLDQASQ 242

RESULT 14
US-09-072-967-355
Sequence 355, Application US/09072967
Patent No. 6592877
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 355:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-072-967-355

Query Match 100.0%; Score 103; DB 4; Length 652;
Best Local Similarity 100.0%; Pred. No. 2e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
DB 223 AETPGCVAYIGISFLDQASQ 242

RESULT 15
US-09-056-556-214
Sequence 214, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 214:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-214

Query Match 100.0%; Score 103; DB 4; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.5e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
DB 289 AETPGCVAYIGISFLDQASQ 309

Search completed: March 10, 2004, 12:14:15
Job time : 13.3077 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:59:11 ; Search time 12.3077 Seconds
(without alignments)
83.892 Million cell updates/sec

Title: US-10-044-703-61

Perfect score: 97

Sequence: 1 NFAGIEAAASAIQGNVTSIH 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/iaa/5A COMB.pcp:*
- 2: /cgn2_6/prodata/2/iaa/5B COMB.pcp:*
- 3: /cgn2_6/prodata/2/iaa/6A COMB.pcp:*
- 4: /cgn2_6/prodata/2/iaa/6B COMB.pcp:*
- 5: /cgn2_6/prodata/2/iaa/PCTUS COMB.pcp:*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	100.0	51	3	US-08-818-112-104
2	97	100.0	51	4	US-08-818-111-99
3	97	100.0	51	4	US-09-056-556-104
4	97	100.0	51	4	US-09-072-596-99
5	97	100.0	51	4	US-09-072-967-104
6	97	100.0	95	2	US-08-465-640-2
7	97	100.0	403	4	US-09-050-739-173
8	97	100.0	404	4	US-09-050-739-172
9	46	47.4	170	2	US-08-477-451-35
10	46	47.4	1732	2	US-08-477-451-14
11	45	46.4	318	4	US-09-252-991A-23917
12	44	45.4	418	4	US-09-198-452A-893
13	44	45.4	418	4	US-09-202-893B-4
14	44	45.4	773	2	US-08-966-389-4
15	44	45.4	773	2	US-09-103-509-4
16	44	45.4	773	2	US-09-102-644-4
17	44	45.4	773	2	US-09-218-032-4
18	43	44.8	739	4	US-09-328-352-6048
19	42	43.3	330	4	US-09-134-000C-6392
20	41	42.3	149	4	US-09-534-238-265
21	41	42.3	282	4	US-09-724-623-113
22	41	42.3	560	4	US-09-252-991A-20141
23	41	42.3	754	4	US-09-392-714-20
24	41	42.3	956	2	US-08-897-443-3
25	40	41.2	395	4	US-08-635-552A-4
26	40	41.2	413	4	US-09-252-991A-29788
27	40	41.2	415	4	US-09-359-268A-25

28	40	41.2	419	2	US-08-270-581-2	Sequence 2, Appli
29	40	41.2	419	4	US-09-146-893-2	Sequence 2, Appli
30	40	41.2	419	4	US-08-675-499A-5	Sequence 5, Appli
31	40	41.2	419	4	US-08-812-008-5	Sequence 5, Appli
32	40	41.2	422	5	PCT-US91-01360-2	Sequence 2, Appli
33	40	41.2	483	4	US-09-489-039A-13018	Sequence 13018, A
34	40	41.2	1014	4	US-09-252-991A-29868	Sequence 29868, A
35	39	40.2	111	4	US-09-134-000C-3792	Sequence 3792, Ap
36	39	40.2	139	4	US-09-134-000C-6763	Sequence 6763, Ap
37	39	40.2	270	4	US-09-252-991A-16844	Sequence 16844, A
38	39	40.2	388	2	US-08-382-505-2	Sequence 2, Appli
39	39	40.2	405	4	US-09-489-039A-7574	Sequence 7574, Ap
40	39	40.2	445	4	US-09-477-135A-135	Sequence 135, App
41	39	40.2	505	3	US-08-993-260-1	Sequence 1, Appli
42	39	40.2	556	4	US-09-252-991A-17793	Sequence 17793, A
43	39	40.2	556	4	US-09-134-000C-4170	Sequence 4170, Ap
44	39	40.2	563	4	US-09-107-532A-4741	Sequence 4741, Ap
45	39	40.2	768	2	US-08-408-519-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-818-112-104
; Sequence 104, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/818,112
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-104

Query Match 100.0%; Score 97; DB 3; Length 51;
Best Local Similarity 100.0%; Pred. No. 4.3e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFAGIEAAASAIQGNVTSIH 20
Db 7 NFAGIEAAASAIQGNVTSIH 26

RESULT 2

US-08-818-111-99
; Sequence 99, Application US/08818111
; Patent No. 6338852

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

TUBERCULOSIS

; NUMBER OF SEQUENCES: 148

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/818,111

; FILING DATE: 13-MAR-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.417C6

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 99:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 51 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-818-111-99

Query Match 100.0%; Score 97; DB 4; Length 51;

Best Local Similarity 100.0%; Pred. No. 4.3e-10;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFAGIEAAASAIQGNVTSIH 20

Db 7 NFAGIEAAASAIQGNVTSIH 26

RESULT 3

US-09-056-556-104

; Sequence 104, Application US/09056556

; Patent No. 6350456

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND

TREAT

; NUMBER OF SEQUENCES: 241

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.457

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 104:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 51 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-09-056-556-104

Query Match 100.0%; Score 97; DB 4; Length 51;

Best Local Similarity 100.0%; Pred. No. 4.3e-10;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFAGIEAAASAIQGNVTSIH 20

Db 7 NFAGIEAAASAIQGNVTSIH 26

RESULT 4

US-09-072-596-99

; Sequence 99, Application US/09072596

; Patent No. 6458366

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Campos-Neto, Antonia

; APPLICANT: Houghton, Raymond

; APPLICANT: Vedwick, Thomas S.

; APPLICANT: Twardzik, Daniel R.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

TUBERCULOSIS

; NUMBER OF SEQUENCES: 350

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/072,596

; FILING DATE: 05-MAY-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.417C9

; TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 99:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-596-99

Query Match 100.0%; Score 97; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 4.3e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAASAIQGNVTSIH 20
DB 7 NFAGIEAASAIQGNVTSIH 26

RESULT 5

US-09-072-967-104
Sequence 104, Application US/09072967
Patent No. 6592877
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-967-104

Query Match 100.0%; Score 97; DB 4; Length 51;
Best Local Similarity 100.0%; Pred. No. 4.3e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAASAIQGNVTSIH 20

DB 7 NFAGIEAASAIQGNVTSIH 26

RESULT 6

US-08-465-640-2
Sequence 2, Application US/08465640
Patent No. 5955077
GENERAL INFORMATION:
APPLICANT: ANDERSEN, Peter
APPLICANT: ANDERSEN, Ase Bengaard
APPLICANT: HASLOV, Kaare
APPLICANT: SORESEN, Anne Lund
TITLE OF INVENTION: TUBERCULOSIS VACCINE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,640
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/123,182
FILING DATE: 20-SEP-1993
PRIOR APPLICATION DATA: PCT/DK94/00273
APPLICATION NUMBER:
FILING DATE: 01-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: COOPER, IVER P
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: ANDERSEN=3A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-465-640-2

Query Match 100.0%; Score 97; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 9.4e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAASAIQGNVTSIH 20

DB 7 NFAGIEAASAIQGNVTSIH 26

RESULT 7

US-09-050-739-173
Sequence 173, Application US/09050739
Patent No. 6641814
GENERAL INFORMATION:
APPLICANT: ANDERSEN, Peter
APPLICANT: NIELSEN, Rikke
APPLICANT: OETTINGER, Thomas
APPLICANT: RASMUSSEN, Peter Birk
APPLICANT: ROSENKRANDS, Ida
APPLICANT: WELLDINGH, Karin
APPLICANT: FLORIO, Walter
TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS

;; TITLE OF INVENTION: DERIVED FROM M. TUBERCULOSIS
;; FILE REFERENCE: 670001-2002.1
;; CURRENT APPLICATION NUMBER: US/09/050,739

;; EARLIER FILING DATE: 1998-03-30
;; EARLIER APPLICATION NUMBER: 0376/97
;; EARLIER FILING DATE: 1997-04-02
;; EARLIER APPLICATION NUMBER: 1277/97
;; EARLIER FILING DATE: 1997-11-10
;; EARLIER APPLICATION NUMBER: 60/044,624
;; EARLIER FILING DATE: 1997-04-18
;; EARLIER APPLICATION NUMBER: 60/070,488
;; EARLIER FILING DATE: 1998-01-05
;; NUMBER OF SEQ ID NOS: 173
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 173
;; LENGTH: 403
;; TYPE: PRT

;; ORGANISM: Mycobacterium tuberculosis
US-09-050-739-173

Query Match 100.0%; Score 97; DB 4; Length 403;
Best Local Similarity 100.0%; Pred. No. 5.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFAGIEAASAIQGNVTSIH 20
Db 28 NFAGIEAASAIQGNVTSIH 47

RESULT 8

US-09-050-739-172
Sequence 172, Application US/09050739
Patent No. 6641814

GENERAL INFORMATION:

APPLICANT: ANDERSEN, Peter
APPLICANT: NIELSEN, Rikke
APPLICANT: OETTINGER, Thomas
APPLICANT: RASMUSSEN, Peter Birk
APPLICANT: ROSENKRANDS, Ida
APPLICANT: WELDRING, Karin
APPLICANT: FLORIO, Walter

;; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
;; FILE OF INVENTION: DERIVED FROM M. TUBERCULOSIS

;; FILE REFERENCE: 670001-2002.1
;; CURRENT APPLICATION NUMBER: US/09/050,739
;; CURRENT FILING DATE: 1998-03-30

;; EARLIER FILING DATE: 1998-03-30
;; EARLIER APPLICATION NUMBER: 0376/97
;; EARLIER FILING DATE: 1997-04-02
;; EARLIER APPLICATION NUMBER: 1277/97
;; EARLIER FILING DATE: 1997-11-10
;; EARLIER APPLICATION NUMBER: 60/044,624
;; EARLIER FILING DATE: 1997-04-18
;; EARLIER APPLICATION NUMBER: 60/070,488
;; EARLIER FILING DATE: 1998-01-05
;; NUMBER OF SEQ ID NOS: 173
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 172
;; LENGTH: 404
;; TYPE: PRT

;; ORGANISM: Mycobacterium tuberculosis
US-09-050-739-172

Query Match 100.0%; Score 97; DB 4; Length 404;
Best Local Similarity 100.0%; Pred. No. 5.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFAGIEAASAIQGNVTSIH 20
Db 316 NFAGIEAASAIQGNVTSIH 335

RESULT 9

US-08-477-451-35

;; Sequence 35, Application US/08477451
;; Patent No. 5928865
;; GENERAL INFORMATION:
;; APPLICANT: Covacci, Antonello
;; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
;; NUMBER OF SEQUENCES: 46
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Chiron Corporation
;; STREET: 4560 Horton Street
;; CITY: Emeryville
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94608-2916
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/477,451
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: McClung, Barbara G.
;; REGISTRATION NUMBER: 33,113
;; REFERENCE/DOCKET NUMBER: 0335.002
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 510-601-2708
;; TELEFAX: 510-655-3542
;; INFORMATION FOR SEQ ID NO: 35:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 170 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-477-451-35

Qy 1 NFAGIEAASAIQGNVTSI 19

Db 14 SWKGIQAALSAIGNVKMI 32

RESULT 10

US-08-477-451-14
Sequence 14, Application US/08477451
Patent No. 5928865

GENERAL INFORMATION:

APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori Cagi Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.

Qy 1 NFAGIEAASAIQGNVTSI 19

Db 14 SWKGIQAALSAIGNVKMI 32

Query Match

Best Local Similarity 47.4%; Score 46; DB 2; Length 170;
Best Local Similarity 52.6%; Pred. No. 1.9;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 NFAGIEAASAIQGNVTSI 19

Db 14 SWKGIQAALSAIGNVKMI 32

RESULT 10

US-08-477-451-14
Sequence 14, Application US/08477451
Patent No. 5928865

GENERAL INFORMATION:

APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori Cagi Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.

Qy 1 NFAGIEAASAIQGNVTSI 19

Db 14 SWKGIQAALSAIGNVKMI 32

Query Match

Best Local Similarity 47.4%; Score 46; DB 2; Length 170;
Best Local Similarity 52.6%; Pred. No. 1.9;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 NFAGIEAASAIQGNVTSI 19

Db 14 SWKGIQAALSAIGNVKMI 32

RESULT 10

US-08-477-451-14
Sequence 14, Application US/08477451
Patent No. 5928865

GENERAL INFORMATION:

APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori Cagi Region
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,451
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.

Qy 1 NFAGIEAASAIQGNVTSI 19

Db 14 SWKGIQAALSAIGNVKMI 32

REGISTRATION NUMBER: 33,113
 REFERENCE/DOCKET NUMBER: 0335.002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 510-601-2708
 TELEFAX: 510-655-3542
 INFORMATION FOR SEQ ID NO. 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1732 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-477-451-14

Query Match 47.4%; Score 46; DB 2; Length 1732;
 Best Local Similarity 52.6%; Pred. No. 36;
 Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NFAGTEAASAIQGNVTSI 19
 DB 654 SWKIQALSALGNVYMI 672

RESULT 11

US-09-252-991A-23917
 Sequence 23917, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 23917
 LENGTH: 318
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-23917

Query Match 46.4%; Score 45; DB 4; Length 318;
 Best Local Similarity 56.2%; Pred. No. 6.3;
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 5 IEAASAIQGNVTSIH 20
 DB 273 IRCALAVQGVTSAH 288

RESULT 12

US-09-198-452A-893
 Sequence 893, Application US/09198452A
 Patent No. 6559294
 GENERAL INFORMATION:
 APPLICANT: Grifffais, R.
 TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
 TITLE OF INVENTION: and treatment of infection
 FILE REFERENCE: 9710-003-999
 CURRENT APPLICATION NUMBER: US/09/198,452A
 CURRENT FILING DATE: 1998-11-24
 NUMBER OF SEQ ID NOS: 6849
 SEQ ID NO 893
 LENGTH: 319
 TYPE: PRT
 ORGANISM: Chlamydia pneumoniae
 US-09-198-452A-893

Query Match 45.4%; Score 44; DB 4; Length 319;
 Best Local Similarity 50.0%; Pred. No. 9.5;
 Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 AGTEAASAIQGNVTSIH 20
 DB 149 ATTEVLASDPQGNVSALH 166

RESULT 13

US-09-202-893B-4
 Sequence 4, Application US/09202893B
 Patent No. 6319692
 GENERAL INFORMATION:
 APPLICANT: KADOTA, Mariko
 APPLICANT: KAWAKI, Mayumi
 APPLICANT: SAWAKI, Saeo
 APPLICANT: SHIRASAWA, Yukiko
 APPLICANT: SONE, Harue
 APPLICANT: SAKO, Tomoyuki
 TITLE OF INVENTION: METHODS FOR TRANSFERRING GENE INTO CHROMOSOME
 FILE REFERENCE: 980794/HG
 CURRENT APPLICATION NUMBER: US/09/202,893B
 CURRENT FILING DATE: 1998-12-22
 PRIOR APPLICATION NUMBER: PCT/JP97/02187
 PRIOR FILING DATE: 1997-06-25
 PRIOR APPLICATION NUMBER: JP 8/184266
 PRIOR FILING DATE: 1996-06-26
 PRIOR APPLICATION NUMBER: JP 8/257764
 PRIOR FILING DATE: 1996-09-06
 NUMBER OF SEQ ID NOS: 9
 SEQ ID NO 4
 LENGTH: 418
 TYPE: PRT
 ORGANISM: Lactobacillus casei
 US-09-202-893B-4

Query Match 45.4%; Score 44; DB 4; Length 418;
 Best Local Similarity 64.3%; Pred. No. 13;
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 FAGTEAASAIQGN 15
 DB 31 FARIKAAAKKVGQN 44

RESULT 14

US-08-966-389-4
 Sequence 4, Application US/08966389
 Patent No. 5843748
 GENERAL INFORMATION:
 APPLICANT: Tetsuya NAKADA
 APPLICANT: Michio KUBOTA
 APPLICANT: Hiroto CHAEN
 APPLICANT: Toshio MIYAKE
 TITLE OF INVENTION: TREHALOSE PHOSPHORYLASE, ITS PREPARATION
 TITLE OF INVENTION: AND USES
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROWDY AND NEIMARK
 STREET: 419 Seventh Street, N.W., Suite 300
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/966,389
 FILING DATE: 07-NOV-1997

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 311,232/1996
; FILING DATE: 8-NOV-1996
; PRIOR APPLICATION DATA: JP 61,716/1997
; APPLICATION NUMBER: JP 61,716/1997
; FILING DATE: 3-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: NAKADA-6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 773 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-966-389-4

Query Match 45.4%; Score 44; DB 2; Length 773;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 NFAGIERAAASAIQGNVTSIH 20
Db 167 NFTGKIRFVSAIDGNVSNIN 186

RESULT 15
US-09-103-509-4
; Sequence 4, Application US/09103509
; Patent No. 5874975
; GENERAL INFORMATION:
; APPLICANT: Tetsuya NAKADA
; APPLICANT: Michio KUBOTA
; APPLICANT: Hiroto CHAEN
; APPLICANT: Toshio MIYAKE
; TITLE OF INVENTION: TREHALOSE PHOSPHORYLASE, ITS PREPARATION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,509
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/966,389
; FILING DATE: 07-NOV-1997
; APPLICATION NUMBER: JP 311,232/1996
; FILING DATE: 8-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 61,716/1997
; FILING DATE: 3-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: NAKADA-6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
```

```
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 773 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-103-509-4

Query Match 45.4%; Score 44; DB 2; Length 773;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 NFAGIERAAASAIQGNVTSIH 20
Db 167 NFTGKIRFVSAIDGNVSNIN 186

Search completed: March 10, 2004, 12:14:16
Job time : 13.3077 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:59:11 ; Search time 12.3077 seconds
(without alignments)
83.892 Million cell updates/sec

Title: US-10-044-703-78

Perfect score: 96

Sequence: 1 IDBLKTNSSLTSLTYHVHV 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96	100.0	163	1	US-08-475-213-5
2	96	100.0	163	5	PCT-US93-11703-1
3	71	74.0	231	2	US-08-997-080-152
4	71	74.0	231	2	US-08-997-362-152
5	71	74.0	231	3	US-09-095-855-152
6	71	74.0	231	3	US-09-324-542-152
7	71	74.0	231	4	US-09-205-426-152
8	70	72.9	228	2	US-08-997-080-147
9	70	72.9	228	2	US-08-997-362-147
10	70	72.9	228	3	US-09-095-855-147
11	70	72.9	228	4	US-09-324-542-147
12	70	72.9	228	4	US-09-205-426-147
13	46	47.9	209	4	US-09-134-001C-5126
14	45	46.9	870	2	US-08-266-311-1
15	45	46.9	870	2	US-08-467-527A-1
16	45	46.9	870	3	US-08-467-528-1
17	44	45.8	885	4	US-09-543-681A-4596
18	44	44.8	782	4	US-09-540-236-2211
19	42	43.8	667	3	US-08-959-004-11
20	42	43.8	2548	4	US-09-172-422-1
21	41.5	43.2	845	1	US-08-356-354-2
22	41.5	43.2	845	2	US-08-778-656-2
23	41.5	43.2	908	1	US-08-356-354-6
24	41.5	43.2	908	2	US-08-778-656-6
25	41.5	43.2	1053	4	US-09-394-272-6
26	41.5	43.2	1054	1	US-08-356-354-4
27	41.5	43.2	1054	2	US-08-778-656-4

28 40 41.7 367 4 US-09-543-681A-4643
29 40 41.7 418 5 PCT-US94-01321-72
30 40 41.7 642 4 US-09-543-681A-4606
31 39 40.6 77 4 US-09-673-395A-412
32 39 40.6 769 4 US-09-788-657-15
33 39 40.6 770 2 US-08-209-521-13
34 39 40.6 770 3 US-08-961-810-123
35 39 40.6 770 3 US-08-352-902D-123
36 39 40.6 770 4 US-09-265-503B-123
37 39 40.6 981 4 US-08-252-931A-16798
38 38.5 40.1 668 4 US-09-697-367-2
39 38.5 40.1 1054 4 US-09-394-272-3
40 38 39.6 121 4 US-09-328-352-5325
41 38 39.6 250 4 US-09-634-238-221
42 38 39.6 264 4 US-09-328-352-7442
43 38 39.6 324 4 US-09-816-095-4
44 38 39.6 328 4 US-09-180-827-7
45 38 39.6 350 4 US-09-489-039A-9711

ALIGNMENTS

RESULT 1

US-08-475-213-5

; Sequence 5, Application US/08475213

; Patent No. 5783674

; GENERAL INFORMATION:

; APPLICANT: Geysen, Hendrik M.

; TITLE OF INVENTION: Method for the use and Synthesis of

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Chiron Corporation

; STREET: 4560 Horton Street

; CITY: Emeryville

; STATE: CA

; COUNTRY: USA

; ZIP: 94608

; COMPUTER READABLE FORM:

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08475,213

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/752,437

; FILING DATE: 06-SEP-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO pct/au90/00062

; FILING DATE: 16-FEB-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: AU PJ2788/89

; FILING DATE: 17-FEB-1989

; ATTORNEY/AGENT INFORMATION:

; NAME: Green, Grant D.

; REGISTRATION NUMBER: 31259

; REFERENCE/DOCKET NUMBER: 0240.002

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 510-601-2706

; TELEFAX: 510-655-3542

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 163 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

```
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium bovis
US-08-475-213-5
Query Match 100.0%; Score 96; DB 1; Length 163;
Best Local Similarity 100.0%; Pred. No. 4.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHVW 20
DB 84 IDELKTNSLLTSILTYHVW 103

RESULT 2
PCT-US93-11703-1
; Sequence 1, Application PC/TUS9311703
; GENERAL INFORMATION:
; APPLICANT: Chiron Mimotopes Pty. Ltd.
; TITLE OF INVENTION: T-Cell Epitopes
; NUMBER OF SEQUENCES: 75
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Grant D. Green
; STREET: 4560 Horton St.
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11703
; FILING DATE: 28-DEC-1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/984,852
; FILING DATE: 02-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Grant D.
; REGISTRATION NUMBER: 31,259
; REFERENCE/DOCKET NUMBER: 0222.101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2706
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 163 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US93-11703-1

Query Match 100.0%; Score 96; DB 5; Length 163;
Best Local Similarity 100.0%; Pred. No. 4.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHVW 20
DB 84 IDELKTNSLLTSILTYHVW 103

RESULT 3
US-08-997-080-152
; Sequence 152, Application US/08997080
; Patent No. 5968524
; GENERAL INFORMATION:
; APPLICANT: WATSON, JAMES D.
; ATTORNEY/AGENT INFORMATION:
; NAME: TAN, PAUL L.J.
```

```
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,080
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 231 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-997-080-152

Query Match 74.0%; Score 71; DB 2; Length 231;
Best Local Similarity 70.0%; Pred. No. 0.0007;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHVW 20
DB 136 LETLKTDSLLTKILTYHVW 155

RESULT 4
US-08-997-362-152
; Sequence 152, Application US/08997362
; Patent No. 5985287
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Hiyama, Jun
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Scott, Linda
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/08/997,362
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
; FILING DATE: June 12, 1997
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
; FILING DATE: August 29, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 231 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-997-362-152

Query Match 74.0%; Score 71; DB 2; Length 231;
Best Local Similarity 70.0%; Pred. No. 0.0007;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTYHVV 20
Db 136 LETLKTDSLLTKILTYHVV 155

RESULT 5
US-09-095-855-152
; Sequence 152, Application US/09095855
; Patent No. 6160093
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Compounds and Methods for
; TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections
; NUMBER OF SEQUENCES: 208
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/095,855
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/705,347
; FILING DATE: 29-AUG-1996
; APPLICATION NUMBER: 08/873,970
; FILING DATE: 12-JUN-1997
; APPLICATION NUMBER: 08/997,362
; FILING DATE: 23-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c3
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 231 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-095-855-152

Query Match 74.0%; Score 71; DB 3; Length 231;
Best Local Similarity 70.0%; Pred. No. 0.0007;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTYHVV 20
Db 136 LETLKTDSLLTKILTYHVV 155

RESULT 6
US-09-324-542-152
; Sequence 152, Application US/09324542
; Patent No. 6328978
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L.J.
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
; FILE REFERENCE: 11000.1007c1
; CURRENT APPLICATION NUMBER: US/09/324,542
; CURRENT FILING DATE: 1999-06-02
; EARLIER APPLICATION NUMBER: US 08/997,080
; EARLIER FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 152:
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
; US-09-324-542-152

Query Match 74.0%; Score 71; DB 4; Length 231;
Best Local Similarity 70.0%; Pred. No. 0.0007;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTYHVV 20
Db 136 LETLKTDSLLTKILTYHVV 155

RESULT 7
US-09-205-426-152
; Sequence 152, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Compounds and Methods for Treatment and
; TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
; FILE REFERENCE: 11000.1002c4
; CURRENT APPLICATION NUMBER: US/09/205,426
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: 09/095,855
; EARLIER FILING DATE: 1998-06-11
; EARLIER APPLICATION NUMBER: 08/997,362
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 08/873,970
; EARLIER FILING DATE: 1997-06-12
; EARLIER APPLICATION NUMBER: 08/705,347
; EARLIER FILING DATE: 1996-08-29
; NUMBER OF SEQ ID NOS: 208

```

SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 152
LENGTH: 231
TYPE: PRT
ORGANISM: Mycobacterium vaccae
US-03-205-426-152

Query Match 74.0%; Score 71; DB 4; Length 231;
Best Local Similarity 70.0%; Pred. No. 0.0007;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 IDELKTNSLSLTSLTYHV 20
DB 136 LETLKTSDLLTKILTYHV 155

RESULT 8

US-08-997-080-147
Sequence 147, Application US/08997080
Patent No. 5968524

GENERAL INFORMATION:
APPLICANT: WATSON, JAMES D.
APPLICANT: TAN, PAUL L.J.
TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
CORRESPONDENCE ADDRESS:
NUMBER OF SEQUENCES: 194

ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,080

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Sleath, Janet

REGISTRATION NUMBER: 37,007

REFERENCE/DOCKET NUMBER: 11000.1007

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-269-0565

TELEFAX: 206-269-0563

TELEX:

INFORMATION FOR SEQ ID NO: 147:

SEQUENCE CHARACTERISTICS:

LENGTH: 228 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-997-080-147

Query Match 72.9%; Score 70; DB 2; Length 228;
Best Local Similarity 65.0%; Pred. No. 0.001;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 IDELKTNSLSLTSLTYHV 20
DB 148 LETLKTSDMLTNILTYHV 167

RESULT 9

US-08-997-362-147
Sequence 147, Application US/08997362
Patent No. 5985287

GENERAL INFORMATION:

APPLICANT: Tan, Paul
APPLICANT: Hiyama, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Scott, Linda

APPLICANT: Prestidge, Ross

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS

NUMBER OF SEQUENCES: 194

CORRESPONDENCE ADDRESS:

ADDRESSEE: Law Offices of Ann W. Speckman

STREET: 2601 Elliott Avenue, Suite 4185

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/997,362

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970

FILING DATE: June 12, 1997

APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347

FILING DATE: August 29, 1996

ATTORNEY/AGENT INFORMATION:

NAME: Sleath, Janet

REGISTRATION NUMBER: 37,007

REFERENCE/DOCKET NUMBER: 11000.1002c2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-269-0565

TELEFAX: 206-269-0563

TELEX:

INFORMATION FOR SEQ ID NO: 147:

SEQUENCE CHARACTERISTICS:

LENGTH: 228 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-997-362-147

Query Match 72.9%; Score 70; DB 2; Length 228;
Best Local Similarity 65.0%; Pred. No. 0.001;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 IDELKTNSLSLTSLTYHV 20
DB 148 LETLKTSDMLTNILTYHV 167

RESULT 10

US-09-095-855-147
Sequence 147, Application US/09095855
Patent No. 6160093

GENERAL INFORMATION:

APPLICANT: Tan, Paul

APPLICANT: Visser, Elizabeth

APPLICANT: Skinner, Margot

APPLICANT: Prestidge, Ross

TITLE OF INVENTION: Compounds and Methods for

TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections

NUMBER OF SEQUENCES: 208

CORRESPONDENCE ADDRESS:

ADDRESSEE: Law Offices of Ann W. Speckman

STREET: 2601 Elliott Avenue, Suite 4185

CITY: Seattle

STATE: WA

```

; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/09/095,855
; APPLICATION NUMBER: US/09/095,855
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/705,347
; FILING DATE: 29-AUG-1996
; APPLICATION NUMBER: 08/873,970
; FILING DATE: 12-JUN-1997
; APPLICATION NUMBER: 08/997,362
; FILING DATE: 23-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 147:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-095-855-147

```

```

Query Match 72.9%; Score 70; DB 3; Length 228;
Best Local Similarity 65.0%; Pred. No. 0.001;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 IDELKTNSLLTSILTYHV 20
: : : : : : : : : : : : : :
Db 148 LETLKTSDMLTNILTYHV 167

```

```

RESULT 11
US-09-324-542-147
; Sequence 147, Application US/09324542
; Patent No. 6328978
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; OF INFECTION: of Immunologically-Mediated Skin Disorders
; FILE REFERENCE: 11000.1007c1
; CURRENT APPLICATION NUMBER: US/09/324,542
; CURRENT FILING DATE: 1999-06-02
; EARLIER APPLICATION NUMBER: US 08/997,080
; EARLIER FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 147
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
; US-09-324-542-147

```

```

Query Match 72.9%; Score 70; DB 4; Length 228;
Best Local Similarity 65.0%; Pred. No. 0.001;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 IDELKTNSLLTSILTYHV 20
: : : : : : : : : : : : : :
Db 148 LETLKTSDMLTNILTYHV 167

```

```

RESULT 12
US-09-205-426-147
; Sequence 147, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Compounds and Methods for Treatment and
; OF INFECTION: Diagnosis of Mycobacterial Infections
; FILE REFERENCE: 11000.1002c4
; CURRENT APPLICATION NUMBER: US/09/205,426
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: 09/095,855
; EARLIER FILING DATE: 1998-06-11
; EARLIER APPLICATION NUMBER: 08/997,362
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 08/873,970
; EARLIER FILING DATE: 1997-06-12
; EARLIER APPLICATION NUMBER: 08/705,347
; EARLIER FILING DATE: 1996-08-29
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 147
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
; US-09-205-426-147

```

```

Query Match 72.9%; Score 70; DB 4; Length 228;
Best Local Similarity 65.0%; Pred. No. 0.001;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 1 IDELKTNSLLTSILTYHV 20
: : : : : : : : : : : : : :
Db 148 LETLKTSDMLTNILTYHV 167

```

```

RESULT 13
US-09-134-001C-5126
; Sequence 5126, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; OF INFECTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5126
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (206)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
; US-09-134-001C-5126

```

```

Query Match 47.9%; Score 46; DB 4; Length 209;
Best Local Similarity 56.2%; Pred. No. 7;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 5 KTNSTNSLLTSILTYHV 20
: : : : : : : : : : : : : :
Db 140 KINASLSREVLTHV 155

```

```
RESULT 14
US-08-266-311-1
; Sequence 1, Application US/08266311
; Patent No. 5814475
; GENERAL INFORMATION:
; APPLICANT: Niepel, Frank
; APPLICANT: Fleckenstein, Bernhard
; TITLE OF INVENTION: Human Herpesvirus Type 6 Protein P100,
; TITLE OF INVENTION: The Corresponding DNA Sequences, Their Preparation and Use
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/266.311
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908.041
; FILING DATE: 06-JUL-1992
; FILING DATE: 06-JUL-1991
; NAME: Einaudi, Carol P.
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 02481-1194-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 870 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-266-311-1
Query Match 46.9%; Score 45; DB 2; Length 870;
Best Local Similarity 50.0%; Pred. No. 51;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTILYHV 20
Db 54 VDDLKTLNLLVLMYHV 73

RESULT 15
US-08-467-527A-1
; Sequence 1, Application US/08467527A
; Patent No. 5827519
; GENERAL INFORMATION:
; APPLICANT: Niepel, Frank
; APPLICANT: Fleckenstein, Bernhard
; TITLE OF INVENTION: Human Herpesvirus Type 6 Protein
; TITLE OF INVENTION: P100, The Corresponding DNA Sequences, Their Preparation
; TITLE OF INVENTION: and Use
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner,
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/266.311
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908.041
; FILING DATE: 06-JUL-1992
; FILING DATE: 06-JUL-1991
; NAME: Einaudi, Carol P.
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 02481-1194-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 870 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-467-527A-1
Query Match 46.9%; Score 45; DB 2; Length 870;
Best Local Similarity 50.0%; Pred. No. 51;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTILYHV 20
Db 54 VDDLKTLNLLVLMYHV 73
```

```
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467.527A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/266.311
; FILING DATE: 27-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/126.435
; FILING DATE: 24-SEP-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908.041
; FILING DATE: 06-JUL-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 91 111 338.9
; FILING DATE: 08-JUL-1991
; NAME: Forman, David S.
; REGISTRATION NUMBER: 33,694
; REFERENCE/DOCKET NUMBER: 02481-1194-04000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 870 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-467-527A-1
Query Match 46.9%; Score 45; DB 2; Length 870;
Best Local Similarity 50.0%; Pred. No. 51;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTILYHV 20
Db 54 VDDLKTLNLLVLMYHV 73

Search completed: March 10, 2004, 12:14:18
Job time : 13.3077 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:59:11 ; Search time 12.3077 Seconds
(without alignments)
83.892 Million cell updates/sec

Title: US-10-044-703-69

Perfect score: 98

Sequence: 1 VSDLKSTAVIPGVVAGQV 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:**

1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/6C-TUS-COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	100.0	159	3	US-08-796-792-2
2	98	100.0	159	4	US-09-491-795-2
3	63	64.3	47	4	US-09-157-689-22
4	62	63.3	47	4	US-09-157-689-21
5	46	46.9	372	4	US-09-489-039A-12516
6	44	44.9	136	3	US-08-673-814-8
7	44	44.9	136	4	US-09-115-824-8
8	44	44.9	298	4	US-09-427-261-1
9	44	44.9	386	4	US-09-489-039A-7410
10	44	44.9	445	2	US-08-630-118A-2
11	44	44.9	445	2	US-08-630-118A-4
12	44	44.9	445	2	US-08-638-399-2
13	44	44.9	445	2	US-08-838-399-4
14	44	44.9	445	2	US-09-003-199-23
15	44	44.9	445	3	US-09-235-839-2
16	44	44.9	445	3	US-09-235-839-4
17	44	44.9	445	4	US-09-327-035-2
18	44	44.9	445	4	US-09-327-035-4
19	44	44.9	445	4	US-09-665-027-6
20	44	44.9	445	1	US-08-349-025-2
21	44	44.9	456	2	US-08-566-096A-2
22	44	44.9	456	2	US-08-668-650B-2
23	44	44.9	456	4	US-09-200-673-2
24	44	44.9	456	4	US-09-194-895-2
25	44	44.9	456	4	US-09-447-907-2
26	44	44.9	456	5	PCT-US95-15646-2
27	44	44.9	456	4	US-09-065-027-8

28 43 43.9 406 4 US-09-252-991A-19757
29 43 43.9 809 4 US-09-489-039A-10903
30 42 42.9 159 4 US-09-252-991A-18201
31 42 42.9 177 4 US-09-107-532A-6228
32 42 42.9 195 4 US-09-134-001C-5224
33 42 42.9 299 4 US-09-904-615-73
34 42 42.9 397 4 US-09-252-991A-27923
35 42 42.9 445 3 US-09-457-046B-73
36 42 42.9 499 2 US-09-032-315-2
37 42 42.9 499 2 US-08-993-318A-2
38 42 42.9 499 3 US-09-399-886-2
39 42 42.9 499 3 US-09-396-260-2
40 42 42.9 499 3 US-09-576-281-2
41 42 42.9 520 1 US-08-482-484-2
42 42 42.9 520 1 US-08-441-147-2
43 42 42.9 520 5 PCT-US95-07536-2
44 42 42.9 738 4 US-09-585-858-6
45 41 41.8 298 4 US-09-345-883-2

ALIGNMENTS

RESULT 1

US-08-796-792-2

; Sequence 2, Application US/08796792

; Patent No. 6087163

; GENERAL INFORMATION:

; APPLICANT: Genmaro, Maria L.

; APPLICANT: Lyashchenko, Konstantin P.

; APPLICANT: Manca, Claudia M.A.

; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS

; TITLE OF INVENTION: SPECIFIC PROTEINS AND GENES, MIXTURES OF ANTIGENS AND USES

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESS: Fish & Richardson, P.C.

; STREET: 45 Rockefeller Plaza, Suite 2800

; CITY: New York

; STATE: NY

; COUNTRY: US

; ZIP: 10111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows95

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/796,792

; FILING DATE: 06-FEB-1997

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/011,364

; FILING DATE: 09-FEB-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Hone, William J.

; REGISTRATION NUMBER: 26,739

; REFERENCE/DOCKET NUMBER: 07763/03301

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-765-5070

; TELEFAX: 212-258-2291

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 159 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

US-08-796-792-2

Query Match 100.0%; Score 98; DB 3; Length 159;
Best Local Similarity 100.0%; Pred. No. 7.2e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	DB	QY	DB
1	56	1	56
VSDLKSTAVIPGYVAGQV 20	VSDLKSTAVIPGYVAGQV 75	VSDLKSTAVIPGYVAGQV 20	VSDLKSTAVIPGYVAGQV 75
<p>RESULT 2</p> <p>US-09-491-795-2</p> <p>Sequence 2, Application US/09491795</p> <p>Patent No. 6596281</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Gennaro, Maria L.</p> <p>APPLICANT: Lyashchenko, Konstantin P.</p> <p>APPLICANT: Manca, Claudia M.A.</p> <p>TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS SPECIFIC PROTEINS AND GENES,</p> <p>TITLE OF INVENTION: MIXTURES OF ANTIGENS AND USES THEREOF</p> <p>FILE REFERENCE: 07763/028002</p> <p>CURRENT APPLICATION NUMBER: US/09/491,795</p> <p>CURRENT FILING DATE: 2000-01-26</p> <p>PRIOR APPLICATION NUMBER: US 08/796,792</p> <p>PRIOR FILING DATE: 1997-02-06</p> <p>PRIOR APPLICATION NUMBER: US 60/011,364</p> <p>PRIOR FILING DATE: 1996-02-09</p> <p>NUMBER OF SEQ ID NOS: 7</p> <p>SOFTWARE: FastSeq for Windows Version 4.0</p> <p>SEQ ID NO 2</p> <p>LENGTH: 159</p> <p>TIPS: PRT</p> <p>ORGANISM: Mycobacterium tuberculosis</p> <p>US-09-491-795-2</p>			
<p>Query Match 100.0%; Score 98; DB 4; Length 159;</p> <p>Best Local Similarity 100.0%; Pred. No. 7.2e-09;</p> <p>Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>			
1	56	1	56
VSDLKSTAVIPGYVAGQV 20	VSDLKSTAVIPGYVAGQV 75	VSDLKSTAVIPGYVAGQV 20	VSDLKSTAVIPGYVAGQV 75
<p>RESULT 3</p> <p>US-09-157-689-22</p> <p>Sequence 22, Application US/09157689</p> <p>Patent No. 6599510</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Horwitz, Marcus A.</p> <p>APPLICANT: Harth, Gunter</p> <p>TITLE OF INVENTION: Abundant Extracellular Products and</p> <p>TITLE OF INVENTION: Methods for Their Production and Use</p> <p>NUMBER OF SEQUENCES: 95</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSES: Kurt A. MacLean</p> <p>STREET: 2029 Century Park East, Suite 3800</p> <p>CITY: Los Angeles</p> <p>STATE: California</p> <p>COUNTRY: U.S.A.</p> <p>ZIP: 90067</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: Patent in Release #1.0, Version #1.30</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/09/157,689</p> <p>FILING DATE:</p> <p>CLASSIFICATION:</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/568,357</p> <p>FILING DATE: 06-DEC-1995</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/551,149</p> <p>FILING DATE: 31-OCT-1995</p> <p>PRIOR APPLICATION DATA:</p>			
<p>Query Match 100.0%; Score 98; DB 4; Length 159;</p> <p>Best Local Similarity 100.0%; Pred. No. 7.2e-09;</p> <p>Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>			
1	56	1	56
VSDLKSTAVIPGYVAGQV 20	VSDLKSTAVIPGYVAGQV 75	VSDLKSTAVIPGYVAGQV 20	VSDLKSTAVIPGYVAGQV 75
<p>RESULT 4</p> <p>US-09-157-689-21</p> <p>Sequence 21, Application US/09157689</p> <p>Patent No. 6599510</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Horwitz, Marcus A.</p> <p>APPLICANT: Harth, Gunter</p> <p>TITLE OF INVENTION: Abundant Extracellular Products and</p> <p>TITLE OF INVENTION: Methods for Their Production and Use</p> <p>NUMBER OF SEQUENCES: 95</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSES: Kurt A. MacLean</p> <p>STREET: 2029 Century Park East, Suite 3800</p> <p>CITY: Los Angeles</p> <p>STATE: California</p> <p>COUNTRY: U.S.A.</p> <p>ZIP: 90067</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: Patent in Release #1.0, Version #1.30</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/09/157,689</p> <p>FILING DATE:</p> <p>CLASSIFICATION:</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/568,357</p> <p>FILING DATE: 06-DEC-1995</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/551,149</p> <p>FILING DATE: 31-OCT-1995</p> <p>PRIOR APPLICATION DATA:</p>			
<p>Query Match 100.0%; Score 98; DB 4; Length 159;</p> <p>Best Local Similarity 100.0%; Pred. No. 7.2e-09;</p> <p>Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>			
1	56	1	56
VSDLKSTAVIPGYVAGQV 20	VSDLKSTAVIPGYVAGQV 75	VSDLKSTAVIPGYVAGQV 20	VSDLKSTAVIPGYVAGQV 75
<p>RESULT 5</p> <p>US-09-157-689-20</p> <p>Sequence 20, Application US/09157689</p> <p>Patent No. 6599510</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Horwitz, Marcus A.</p> <p>APPLICANT: Harth, Gunter</p> <p>TITLE OF INVENTION: Abundant Extracellular Products and</p> <p>TITLE OF INVENTION: Methods for Their Production and Use</p> <p>NUMBER OF SEQUENCES: 95</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSES: Kurt A. MacLean</p> <p>STREET: 2029 Century Park East, Suite 3800</p> <p>CITY: Los Angeles</p> <p>STATE: California</p> <p>COUNTRY: U.S.A.</p> <p>ZIP: 90067</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: Patent in Release #1.0, Version #1.30</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/09/157,689</p> <p>FILING DATE:</p> <p>CLASSIFICATION:</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/568,357</p> <p>FILING DATE: 06-DEC-1995</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/551,149</p> <p>FILING DATE: 31-OCT-1995</p> <p>PRIOR APPLICATION DATA:</p>			
<p>Query Match 100.0%; Score 98; DB 4; Length 159;</p> <p>Best Local Similarity 100.0%; Pred. No. 7.2e-09;</p> <p>Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>			
1	56	1	56
VSDLKSTAVIPGYVAGQV 20	VSDLKSTAVIPGYVAGQV 75	VSDLKSTAVIPGYVAGQV 20	VSDLKSTAVIPGYVAGQV 75
<p>RESULT 6</p> <p>US-09-157-689-19</p> <p>Sequence 19, Application US/09157689</p> <p>Patent No. 6599510</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Horwitz, Marcus A.</p> <p>APPLICANT: Harth, Gunter</p> <p>TITLE OF INVENTION: Abundant Extracellular Products and</p> <p>TITLE OF INVENTION: Methods for Their Production and Use</p> <p>NUMBER OF SEQUENCES: 95</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSES: Kurt A. MacLean</p> <p>STREET: 2029 Century Park East, Suite 3800</p> <p>CITY: Los Angeles</p> <p>STATE: California</p> <p>COUNTRY: U.S.A.</p> <p>ZIP: 90067</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: Patent in Release #1.0, Version #1.30</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/09/157,689</p> <p>FILING DATE:</p> <p>CLASSIFICATION:</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/568,357</p> <p>FILING DATE: 06-DEC-1995</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/551,149</p> <p>FILING DATE: 31-OCT-1995</p> <p>PRIOR APPLICATION DATA:</p>			
<p>Query Match 100.0%; Score 98; DB 4; Length 159;</p> <p>Best Local Similarity 100.0%; Pred. No. 7.2e-09;</p> <p>Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>			
1	56	1	56
VSDLKSTAVIPGYVAGQV 20	VSDLKSTAVIPGYVAGQV 75	VSDLKSTAVIPGYVAGQV 20	VSDLKSTAVIPGYVAGQV 75
<p>RESULT 7</p> <p>US-09-157-689-18</p> <p>Sequence 18, Application US/09157689</p> <p>Patent No. 6599510</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Horwitz, Marcus A.</p> <p>APPLICANT: Harth, Gunter</p> <p>TITLE OF INVENTION: Abundant Extracellular Products and</p> <p>TITLE OF INVENTION: Methods for Their Production and Use</p> <p>NUMBER OF SEQUENCES: 95</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSES: Kurt A. MacLean</p> <p>STREET: 2029 Century Park East, Suite 3800</p> <p>CITY: Los Angeles</p> <p>STATE: California</p> <p>COUNTRY: U.S.A.</p> <p>ZIP: 90067</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: Patent in Release #1.0, Version #1.30</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/09/157,689</p> <p>FILING DATE:</p> <p>CLASSIFICATION:</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 08/568,357</</p>			

ATTORNEY/AGENT INFORMATION:
NAME: Whitham, Michael E.
REFERENCE/DOCKET NUMBER: VVIP 95-067
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-391-2510
TELEFAX: 703-391-9035
TELEX: 283072
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 136 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-115-824-8

Query Match 44.9%; Score 44; DB 4; Length 136;
Best Local Similarity 58.8%; Pred. No. 7.3;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 VSDLKSTAVIPGTPVA 17
DB 65 VSDLNIIAFVITGPEMA 81

RESULT 8
US-09-427-261-1
Sequence 1, Application US/09427261A
Patent No. 6368811
GENERAL INFORMATION:
APPLICANT: Grootjans, Jan
APPLICANT: Zimmerman, Pascale
APPLICANT: David, Guido
TITLE OF INVENTION: Syndecan Interacting Proteins and the Use Thereof
FILE REFERENCE: 2676-4206US
CURRENT APPLICATION NUMBER: US/09/427,261A
CURRENT FILING DATE: 1999-10-25
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 298
TYPE: PRT
ORGANISM: Unknown Organism
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: Syndecan Proteins
NAME/KEY: VARIANT
LOCATION: (119)-(214)
OTHER INFORMATION: Positions 119 and 214 can be K or R
US-09-427-261-1

Query Match 44.9%; Score 44; DB 4; Length 298;
Best Local Similarity 38.9%; Pred. No. 19;
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 3 DLKSTAVIPGTPVAGQV 20
DB 65 EIRASVAVVSGAPLQGL 82

RESULT 9
US-09-489-039A-7410
Sequence 7410, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29

NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 7410
LENGTH: 386
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7410

Query Match 44.9%; Score 44; DB 4; Length 386;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 VSDLKSTAVIPGYPV 16
DB 279 VSDVQGEISWFFGYPV 294

RESULT 10
US-08-630-118A-2
Sequence 2, Application US/08630118A
Patent No. 5919901
GENERAL INFORMATION:
APPLICANT: Hu Ph.D., Yinghe
APPLICANT: McCaleb Ph.D., Michael L.
APPLICANT: Bloomquist Ph.D., Brian T.
APPLICANT: Flores-Riveros Ph.D., Jaime R.
APPLICANT: Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
TITLE OF INVENTION: Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boshnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/630,118A
APPLICATION NUMBER: US/08/630,118A
FILING DATE: April 8, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,142
REFERENCE/DOCKET NUMBER: 96,149/WH 405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 715-1000
TELEFAX: (312) 715-1234
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 445 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-630-118A-2

Query Match 44.9%; Score 44; DB 2; Length 445;
Best Local Similarity 47.4%; Pred. No. 30;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 SLDKSTAVIPGYPVAGQV 20
DB 320 SQLSPSSKVPICFICFV 338

RESULT 11
US-08-630-118A-4
Sequence 4, Application US/08630118A
Patent No. 5919901

```

; GENERAL INFORMATION:
; APPLICANT: Hu Ph.D., Yinghe
; APPLICANT: McCaleb Ph.D., Michael L.
; APPLICANT: Bloomquist Ph.D., Brian T.
; APPLICANT: Flores-Riveros Ph.D., Jaime R.
; APPLICANT: Cornfield Ph.D., Linda J.
; TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
; TITLE OF INVENTION: Sequences
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,118A
; FILING DATE: April 8, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenfield Ph.D., Michael S.
; REGISTRATION NUMBER: 37,142
; REFERENCE/DOCKET NUMBER: 96,149/WH 405
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)715-1000
; TELEFAX: (312)715-1234
; INFORMATION FOR SEQ ID NO: 4:
; LENGTH: 445 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-630-118A-4

Query Match 44.9%; Score 44; DB 2; Length 445;
Best Local Similarity 47.4%; Pred. No. 30;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 SDLSSTAVIPGVPGVQV 20
Db 320 SQLSPSSKVIPIGVPICFV 338

RESULT 12
US-08-838-399-2
; Sequence 2, Application US/08838399
; Patent No. 5965392
; GENERAL INFORMATION:
; APPLICANT: Hu Ph.D., Yinghe
; APPLICANT: McCaleb Ph.D., Michael L.
; APPLICANT: Bloomquist Ph.D., Brian T.
; APPLICANT: Flores-Riveros Ph.D., Jaime R.
; APPLICANT: Cornfield Ph.D., Linda J.
; TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
; TITLE OF INVENTION: Sequences
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/838,399
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenfield Ph.D., Michael S.
; REGISTRATION NUMBER: 37,147
; REFERENCE/DOCKET NUMBER: 96,149/WH 405
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)715-1000
; TELEFAX: (312)715-1234
; INFORMATION FOR SEQ ID NO: 4:
; LENGTH: 445 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-838-399-4

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/838,399
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenfield Ph.D., Michael S.
; REGISTRATION NUMBER: 37,147
; REFERENCE/DOCKET NUMBER: 96,149/WH 405
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)715-1000
; TELEFAX: (312)715-1234
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-838-399-2

Query Match 44.9%; Score 44; DB 2; Length 445;
Best Local Similarity 47.4%; Pred. No. 30;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 SDLSSTAVIPGVPGVQV 20
Db 320 SQLSPSSKVIPIGVPICFV 338

RESULT 13
US-08-838-399-4
; Sequence 4, Application US/08838399
; Patent No. 5965392
; GENERAL INFORMATION:
; APPLICANT: Hu Ph.D., Yinghe
; APPLICANT: McCaleb Ph.D., Michael L.
; APPLICANT: Bloomquist Ph.D., Brian T.
; APPLICANT: Flores-Riveros Ph.D., Jaime R.
; APPLICANT: Cornfield Ph.D., Linda J.
; TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
; TITLE OF INVENTION: Sequences
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/838,399
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenfield Ph.D., Michael S.
; REGISTRATION NUMBER: 37,147
; REFERENCE/DOCKET NUMBER: 96,149/WH 405
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)715-1000
; TELEFAX: (312)715-1234
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-838-399-4

```

Query Match 44.9%; Score 44; DB 2; Length 445;
Best Local Similarity 47.4%; Pred. No. 30;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 SDLXSSTAVIPGYPVAGQV 20
DB 320 SQLSPSSKVIPIGVICFEV 338

RESULT 14

US-09-003-199-23
; Sequence 23, Application US/09003199
; Patent No. 5985616
; GENERAL INFORMATION:
; APPLICANT: Parker, Eric M
; APPLICANT: Strader, Catherine D
; APPLICANT: Rudinski, Mark S
; TITLE OF INVENTION: CHIMERIC MAMMALIAN NPX Y5 RECEPTORS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corporation
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: NJ
; COUNTRY: USA
; ZIP: 07033-0530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.5.3
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/003,199
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Thampoe, Immac J.
; REGISTRATION NUMBER: 36,322
; REFERENCE/DOCKET NUMBER: CN0775
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)298-5061
; TELEFAX: (908)298-5388
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-003-199-23

Query Match 44.9%; Score 44; DB 2; Length 445;
Best Local Similarity 47.4%; Pred. No. 30;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 SDLXSSTAVIPGYPVAGQV 20
DB 320 SQLSPSSKVIPIGVICFEV 338

RESULT 15

US-09-235-839-2
; Sequence 2, Application US/09235839
; Patent No. 6207799
; GENERAL INFORMATION:
; APPLICANT: Hu Ph.D., Yinghe
; APPLICANT: McCaleb Ph.D., Michael L.
; APPLICANT: Bloomquist Ph.D., Brian T.
; APPLICANT: Flores-Riveros Ph.D., Jaime R.
; APPLICANT: Cornfield Ph.D., Linda J.
; TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,839
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/630,118
; FILING DATE: April 8, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenfield Ph.D., Michael S.
; REGISTRATION NUMBER: 37,142
; REFERENCE/DOCKET NUMBER: 96,149-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)913-0001
; TELEFAX: (312)913-0002
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-235-839-2

Query Match 44.9%; Score 44; DB 3; Length 445;
Best Local Similarity 47.4%; Pred. No. 30;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 SDLXSSTAVIPGYPVAGQV 20
DB 320 SQLSPSSKVIPIGVICFEV 338

Search completed: March 10, 2004, 12:14:16
Job time: 12.3077 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:59:11 ; Search time 11.6923 Seconds
(without alignments)
83.892 Million cell updates/sec

Title: US-10-044-703-80

Perfect score: 95

Sequence: 1 GVSTANATVMIDSLVMP 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pcp.*
- 6: /cgn2_6/ptodata/2/iaa/backfile1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95	100.0	163	1	US-08-475-213-5
2	95	100.0	163	5	PCT-US93-11703-1
3	85	89.5	228	2	US-08-997-080-147
4	85	89.5	228	3	US-08-997-362-147
5	85	89.5	228	3	US-09-095-855-147
6	85	89.5	228	4	US-09-324-542-147
7	85	89.5	228	4	US-09-405-426-147
8	85	89.5	231	2	US-08-997-080-152
9	85	89.5	231	3	US-09-095-855-152
10	85	89.5	231	4	US-09-324-542-152
11	85	89.5	231	4	US-09-405-426-152
12	85	89.5	231	1	US-08-475-213-8
13	59	62.1	12	1	US-08-751-189-3
14	45	47.4	2627	2	US-09-060-836-3
15	45	47.4	2627	3	US-09-184-445-3
16	44	46.3	18	1	US-08-426-627-13
17	44	46.3	18	1	US-08-426-627-2
18	44	46.3	811	1	US-08-426-627-22
19	44	46.3	811	1	US-08-426-627-15
20	43	45.3	135	1	US-08-477-396A-2
21	43	45.3	212	2	US-08-477-396A-4
22	43	45.3	777	2	US-08-426-627-4
23	43	45.3	779	1	US-08-426-627-24
24	43	45.3	779	1	US-08-426-627-6
25	43	45.3	836	1	US-08-426-627-23
26	43	45.3	837	1	US-08-233-788A-42
27	42	44.2	814	1	US-08-233-788A-42

28	41	43.2	324	4	US-09-252-991A-21941	Sequence 21941, A
29	41	43.2	393	4	US-09-328-352-7939	Sequence 7939, Ap
30	41	43.2	683	1	US-07-878-960-2	Sequence 2, Appli
31	41	43.2	683	2	US-08-477-396A-17	Sequence 17, Appl
32	41	43.2	683	4	US-09-401-064-122	Sequence 122, App
33	40	42.1	590	3	US-09-413-814-89	Sequence 89, Appl
34	40	42.1	591	3	US-09-413-814-76	Sequence 76, Appl
35	39	41.1	136	1	US-08-426-627-17	Sequence 17, Appl
36	39	41.1	137	4	US-09-252-991A-32243	Sequence 32243, A
37	39	41.1	298	4	US-09-489-039A-7869	Sequence 7869, Ap
38	39	41.1	318	1	US-08-220-958-4	Sequence 4, Appli
39	39	41.1	488	4	US-09-252-991A-30531	Sequence 30531, A
40	39	41.1	527	4	US-09-540-236-2518	Sequence 2518, Ap
41	39	41.1	824	4	US-09-252-991A-32329	Sequence 32329, A
42	39	41.1	1111	1	US-08-317-450B-15	Sequence 15, Appl
43	39	41.1	1111	3	US-08-800-593-15	Sequence 15, Appl
44	39	41.1	1112	4	US-09-717-364A-27	Sequence 27, Appl
45	39	41.1	1193	1	US-08-317-450B-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-08-475-213-5
; Sequence 5, Application US/08475213
; Patent No. 5783674
; GENERAL INFORMATION:
; APPLICANT: Geyesen, Hendrik M.
; TITLE OF INVENTION: Method for the use and Synthesis of
; TITLE OF INVENTION: Peptides
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/475,213
APPLICATION NUMBER: US/08/475,213
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/752,437
FILING DATE: 06-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO pct/au90/00062
FILING DATE: 16-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ2788/89
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Green, Grant D.
REGISTRATION NUMBER: 31259
REFERENCE/DOCKET NUMBER: 0240.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2706
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 163 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO

```
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium bovis
US-08-475-213-5

Query Match      100.0%; Score 95; DB 1; Length 163;
Best Local Similarity 100.0%; Pred. No. 3.4e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GVSTANATVYMIDSVLMPP 19
Db      144 GVSTANATVYMIDSVLMPP 162

RESULT 2
PCT-US93-11703-1
; Sequence 1, Application PC/TUS9311703
; GENERAL INFORMATION:
; APPLICANT: Chiron Mimotopes Pty. Ltd.
; TITLE OF INVENTION: T-Cell Epitopes
; NUMBER OF SEQUENCES: 75
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Grant D. Green
; STREET: 4560 Horton St.
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11703
; FILING DATE: 28-DEC-1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/984,852
; FILING DATE: 02-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Grant D.
; REGISTRATION NUMBER: 31,259
; REFERENCE/DOCKET NUMBER: 0222.101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2706
; TELEFAX: 510-855-3542
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 163 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US93-11703-1

Query Match      100.0%; Score 95; DB 5; Length 163;
Best Local Similarity 100.0%; Pred. No. 3.4e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GVSTANATVYMIDSVLMPP 19
Db      144 GVSTANATVYMIDSVLMPP 162

RESULT 3
US-08-997-080-147
; Sequence 147, Application US/08997080
; Patent No. 5968524
; GENERAL INFORMATION:
; APPLICANT: WATSON, JAMES D.
; APPLICANT: TAN, PAUL L.J.
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,080
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 147:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-997-080-147

Query Match      89.5%; Score 85; DB 2; Length 228;
Best Local Similarity 84.2%; Pred. No. 3e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GVSTANATVYMIDSVLMPP 19
Db      208 GVQTANATVYLIDTVLMPP 226

RESULT 4
US-08-997-362-147
; Sequence 147, Application US/08997362
; Patent No. 5985287
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Hiyama, Jun
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Scott, Linda
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
```



```
; APPLICATION NUMBER: US/08/997,362
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
; FILING DATE: June 12, 1997
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
; FILING DATE: August 29, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 147:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-997-362-147
Query Match 89.5%; Score 85; DB 2; Length 228;
Best Local Similarity 84.2%; Pred. No. 3e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
Db 208 GVQTANATVYLIDTVLMPP 226

RESULT 5
US-09-095-855-147
; Sequence 147, Application US/09095855
; Patent No. 6160093
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Prestidge, Rose
; TITLE OF INVENTION: Compounds and Methods for
; TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections
; NUMBER OF SEQUENCES: 208
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/095,855
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/705,347
; FILING DATE: 29-AUG-1996
; APPLICATION NUMBER: 08/873,970
; FILING DATE: 12-JUN-1997
; APPLICATION NUMBER: 08/997,362
; FILING DATE: 23-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c3
; TELECOMMUNICATION INFORMATION:
```

```
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 147:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-095-855-147
Query Match 89.5%; Score 85; DB 3; Length 228;
Best Local Similarity 84.2%; Pred. No. 3e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
Db 208 GVQTANATVYLIDTVLMPP 226

RESULT 6
US-09-324-542-147
; Sequence 147, Application US/09324542
; Patent No. 6328978
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L.J.
; APPLICANT: Prestidge, Rose
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
; FILE REFERENCE: 11000.1007c1
; CURRENT APPLICATION NUMBER: US/09/324,542
; CURRENT FILING DATE: 1999-06-02
; EARLIER APPLICATION NUMBER: US 08/997,080
; EARLIER FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 147
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
;
US-09-324-542-147
Query Match 89.5%; Score 85; DB 4; Length 228;
Best Local Similarity 84.2%; Pred. No. 3e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
Db 208 GVQTANATVYLIDTVLMPP 226

RESULT 7
US-09-205-426-147
; Sequence 147, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Compounds and Methods for Treatment and
; TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
; FILE REFERENCE: 11000.1002c4
; CURRENT APPLICATION NUMBER: US/09/205,426
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: 09/095,855
; EARLIER FILING DATE: 1998-06-11
; EARLIER APPLICATION NUMBER: 08/997,362
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 08/873,970
; EARLIER FILING DATE: 1997-06-12
; EARLIER APPLICATION NUMBER: 08/705,347
; EARLIER FILING DATE: 1996-08-29
; NUMBER OF SEQ ID NOS: 208
```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 147
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-09-205-426-147

Query Match 89.5%; Score 85; DB 4; Length 228;
Best Local Similarity 84.2%; Pred. No. 3e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMP 19
DB 208 GVQTANATVYLIDTVLMP 226

RESULT 8
US-08-997-080-152
; Sequence 152, Application US/08997080
; Patent No. 5968524
; GENERAL INFORMATION:
; APPLICANT: WATSON, JAMES D.
; APPLICANT: TAN, PAUL L.J.
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,080
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1007
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:

INFORMATION FOR SEQ ID NO: 152:
SEQUENCE CHARACTERISTICS:
LENGTH: 231 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-997-080-152

Query Match 89.5%; Score 85; DB 2; Length 231;
Best Local Similarity 84.2%; Pred. No. 3.1e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMP 19
DB 196 GVQTANATVYLIDTVLMP 214

RESULT 9
US-08-997-362-152
; Sequence 152, Application US/08997362
; Patent No. 5985287

; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Hiyama, Jun
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Scott, Linda
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,362
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
FILING DATE: June 12, 1997
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
FILING DATE: August 29, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:

INFORMATION FOR SEQ ID NO: 152:
SEQUENCE CHARACTERISTICS:
LENGTH: 231 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-997-362-152

Query Match 89.5%; Score 85; DB 2; Length 231;
Best Local Similarity 84.2%; Pred. No. 3.1e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMP 19
DB 196 GVQTANATVYLIDTVLMP 214

RESULT 10
US-09-095-855-152
; Sequence 152, Application US/09095855
; Patent No. 6160093
; GENERAL INFORMATION:

APPLICANT: Tan, Paul
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Compounds and Methods for
TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections
NUMBER OF SEQUENCES: 208
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA

```

; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION NUMBER: US/09/095,855
; APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/705,347
; FILING DATE: 29-AUG-1996
; APPLICATION NUMBER: 08/873,970
; FILING DATE: 12-JUN-1997
; APPLICATION NUMBER: 08/997,362
; FILING DATE: 23-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 152:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 231 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-095-855-152

```

```

Query Match      89.5%; Score 85; DB 3; Length 231;
Best Local Similarity 84.2%; Pred. No. 3.1e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1  GVSTANATVYMIDSVLMPP 19
DB      196  GVQTANATVYLIDTVLMPP 214

```

```

RESULT 11
US-09-324-542-152
; Sequence 152, Application US/09324542
; Patent No. 6328978
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L.J.
; TITLE OF INVENTION: Prestidge, Ross
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
; FILE REFERENCE: 11000.1007c1
; CURRENT APPLICATION NUMBER: US/09/324,542
; CURRENT FILING DATE: 1999-06-02
; EARLIER APPLICATION NUMBER: US 08/997,080
; EARLIER FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 152
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
; US-09-324-542-152

```

```

Query Match      89.5%; Score 85; DB 4; Length 231;
Best Local Similarity 84.2%; Pred. No. 3.1e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1  GVSTANATVYMIDSVLMPP 19
DB      196  GVQTANATVYLIDTVLMPP 214

```

```

RESULT 12
US-09-205-426-152
; Sequence 152, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L.J.
; TITLE OF INVENTION: Compounds and Methods for Treatment and
; TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
; FILE REFERENCE: 11000.1002c4
; CURRENT APPLICATION NUMBER: US/09/205,426
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: 09/095,855
; EARLIER FILING DATE: 1998-06-11
; EARLIER APPLICATION NUMBER: 08/997,362
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 08/873,970
; EARLIER FILING DATE: 1997-06-12
; EARLIER APPLICATION NUMBER: 08/705,347
; EARLIER FILING DATE: 1996-08-29
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 152
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
; US-09-205-426-152

```

```

Query Match      89.5%; Score 85; DB 4; Length 231;
Best Local Similarity 84.2%; Pred. No. 3.1e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1  GVSTANATVYMIDSVLMPP 19
DB      196  GVQTANATVYLIDTVLMPP 214

```

```

RESULT 13
US-08-475-213-8
; Sequence 8, Application US/08475213
; Patent No. 5783674
; GENERAL INFORMATION:
; APPLICANT: Geysen, Hendrik M.
; TITLE OF INVENTION: Method for the use and Synthesis of
; TITLE OF INVENTION: Peptides
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,213
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/752,437
; FILING DATE: 06-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO pct/au90/00062
; FILING DATE: 16-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ2788/89
; FILING DATE: 17-FEB-1989

```

ATTORNEY/AGENT INFORMATION:
NAME: Green, Grant D.
REGISTRATION NUMBER: 31259
REFERENCE/DOCKET NUMBER: 0240.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2706
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-475-213-8

Query Match 62.1%; Score 59; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00033;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ATVMIDSVLMP 18
DB 1 ATVMIDSVLMP 12

RESULT 14
US-08-751-189-3
Sequence 3, Application US/08751189
Patent No. 5919656
GENERAL INFORMATION:
APPLICANT: Harrington, Lea A.
TITLE OF INVENTION: No. 5919656el Genes Encoding Telomerase Protein
TITLE OF INVENTION: 1
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen, Inc.
STREET: 1840 De Havilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/751,189
FILING DATE: 15-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy A.
REGISTRATION NUMBER: 34,688
REFERENCE/DOCKET NUMBER: A-433
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2627 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-751-189-3

Query Match 47.4%; Score 45; DB 2; Length 2627;
Best Local Similarity 66.7%; Pred. No. 73;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 VSTANATVYMD 13
DB 1693 VGTANGTVYILD 1704

RESULT 15
US-09-060-836-3
Sequence 3, Application US/09060836
Patent No. 5981707
GENERAL INFORMATION:
APPLICANT: Harrington, Lea A.
TITLE OF INVENTION: No. 5981707el Genes Encoding Telomerase Protein
TITLE OF INVENTION: 1
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen, Inc.
STREET: 1840 De Havilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/060,836
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/751,189
FILING DATE: 15-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy A.
REGISTRATION NUMBER: 34,688
REFERENCE/DOCKET NUMBER: A-433
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2627 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-09-060-836-3

Query Match 47.4%; Score 45; DB 2; Length 2627;
Best Local Similarity 66.7%; Pred. No. 73;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 VSTANATVYMD 13
DB 1693 VGTANGTVYILD 1704

Search completed: March 10, 2004, 12:14:18
Job time : 11.6923 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 12:11:07 ; Search time 25.1282 Seconds
(without alignments)
168.061 Million cell updates/sec

Title: US-10-044-703-47

Perfect score: 96

Sequence: 1 NFLLPDAQSIQAAAAGFASK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09D_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	20	9 US-09-813-333-47	Sequence 47, Appl
2	96	100.0	20	13 US-10-044-703-47	Sequence 47, Appl
3	96	100.0	374	9 US-09-287-849-6	Sequence 6, Appli
4	96	100.0	374	9 US-09-287-849-40	Sequence 40, Appl
5	96	100.0	374	14 US-10-193-002-148	Sequence 148, App
6	96	100.0	374	14 US-10-193-002-150	Sequence 150, App
7	96	100.0	374	14 US-10-084-843-153	Sequence 153, App
8	96	100.0	374	14 US-10-084-843-155	Sequence 155, App
9	96	100.0	374	14 US-10-359-460-6	Sequence 6, Appli
10	96	100.0	374	14 US-10-359-460-40	Sequence 40, Appl
11	96	100.0	374	14 US-10-098-732A-39	Sequence 39, Appl
12	96	100.0	652	14 US-10-193-002-350	Sequence 350, App
13	96	100.0	652	14 US-10-084-843-355	Sequence 355, App
14	96	100.0	802	9 US-09-287-849-10	Sequence 10, Appl
15	96	100.0	802	14 US-10-193-002-209	Sequence 209, App

Sequence 346, App
Sequence 214, App
Sequence 351, App
Sequence 10, Appl
Sequence 58, Appl
Sequence 58, Appl
Sequence 42, Appl
Sequence 81, Appl
Sequence 82, Appl
Sequence 5, Appli
Sequence 2931, Ap
Sequence 14562, A
Sequence 1730, A
Sequence 9046, Ap
Sequence 52, Appl
Sequence 2, Appli
Sequence 19, Appl
Sequence 1865, Ap
Sequence 7624, Ap
Sequence 5066, Ap
Sequence 13876, A
Sequence 41, Appl
Sequence 5941, Ap
Sequence 1119, Ap
Sequence 252, App
Sequence 4, Appli
Sequence 27, Appl
Sequence 84, Appl

ALIGNMENTS

RESULT 1

US-09-813-333-47
; Sequence 47, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:
; APPLICANT: DeGroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 47
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-813-333-47

Query Match 100.0%; Score 96; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. NO. 6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1 NFLLPDAQSIQAAAAGFASK 20
Db	1 NFLLPDAQSIQAAAAGFASK 20

RESULT 2

US-10-044-703-47
; Sequence 47, Application US/10044703
; Publication No. US2002019233A1
; GENERAL INFORMATION:
; APPLICANT: DeGroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703

```

; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 47
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-47

```

```

Query Match      100.0%; Score 96; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 NFFLPDAQSIQAAAAAGFASK 20
    |||||
Db 1 NFFLPDAQSIQAAAAAGFASK 20
    |||||

```

RESULT 3

```

US-09-287-849-6
; Sequence 6, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

```

```

; OTHER INFORMATION: Description of Artificial Sequence: tti-fusion
US-09-287-849-6

```

```

Query Match      100.0%; Score 96; DB 9; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 NFFLPDAQSIQAAAAAGFASK 20
    |||||
Db 272 NFFLPDAQSIQAAAAAGFASK 291
    |||||

```

RESULT 4

```

US-09-287-849-40
; Sequence 40, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.

```

```

; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 40
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: 38 kD antigen
US-09-287-849-40

```

```

Query Match      100.0%; Score 96; DB 9; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 NFFLPDAQSIQAAAAAGFASK 20
    |||||
Db 272 NFFLPDAQSIQAAAAAGFASK 291
    |||||

```

RESULT 5

```

US-10-193-002-148
; Sequence 148, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.

```

```

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

```

```

; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998

```

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 148:
US-10-193-002-148
Query Match 100.0%; Score 96; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFFLPDAQSIQAAAAGFASK 20
Db 272 NFFLPDAQSIQAAAAGFASK 291

RESULT 6
US-10-193-002-150
; Sequence 150, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedwick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 150:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>

```

```

;
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 150:
US-10-193-002-150
Query Match 100.0%; Score 96; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFFLPDAQSIQAAAAGFASK 20
Db 272 NFFLPDAQSIQAAAAGFASK 291

RESULT 7
US-10-084-843-153
; Sequence 153, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedwick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 153:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 153:
US-10-084-843-153
Query Match 100.0%; Score 96; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFFLPDAQSIQAAAAGFASK 20
Db 272 NFFLPDAQSIQAAAAGFASK 291

```

RESULT 8
US-10-084-843-155
; Sequence 155, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, David C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 155:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; LENGTH: 374 amino acids
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 155:
US-10-084-843-155
Query Match 100.0%; Score 96; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 NFLLPDAQSIQAAAAGFASK 20
Db 272 NFLLPDAQSIQAAAAGFASK 291
RESULT 9
US-10-084-843-155
; Sequence 6, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 40
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: 38 kD antigen
US-10-359-460-40

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-460-6
Query Match 100.0%; Score 96; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 NFLLPDAQSIQAAAAGFASK 20
Db 272 NFLLPDAQSIQAAAAGFASK 291
RESULT 10
US-10-359-460-40
; Sequence 40, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 40
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: 38 kD antigen
US-10-359-460-40


```
Query Match      100.0%; Score 96; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 NFFLPDAQSIQAAAAGPASK 20
DB      272 NFFLPDAQSIQAAAAGPASK 291

RESULT 11
US-10-098-732A-39
; Sequence 39, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-0120100S
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 39
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: 38 kD
US-10-098-732A-39

Query Match      100.0%; Score 96; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 NFFLPDAQSIQAAAAGPASK 20
DB      272 NFFLPDAQSIQAAAAGPASK 291

RESULT 12
US-10-193-002-350
; Sequence 350, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392

US-10-084-843-355
; Sequence 355, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
```

REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 355:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 355:
US-10-084-843-355

Query Match 100.0%; Score 96; DB 14; Length 652;
Best Local Similarity 100.0%; Pred. No. 3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFLLPDAQSIQAAAAGFASK 20
|||||
Db 256 NFLLPDAQSIQAAAAGFASK 275

RESULT 14

US-09-287-849-10
Sequence 10, Application US/09287849
Patent No. US20020009459A1
GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/09/287,849
PRIOR FILING DATE: 1999-04-07
PRIOR FILING DATE: 1997-03-13
PRIOR FILING DATE: 1997-10-01
PRIOR FILING DATE: 1998-02-18
PRIOR FILING DATE: 1998-04-07
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 802
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: tetra-fusion
US-09-287-849-10

Query Match 100.0%; Score 96; DB 9; Length 802;
Best Local Similarity 100.0%; Pred. No. 3.7e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFLLPDAQSIQAAAAGFASK 20
|||||
Db 322 NFLLPDAQSIQAAAAGFASK 341

RESULT 15

US-10-193-002-209
Sequence 209, Application US/10193002
Publication No. US20030135026A1
GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS

NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 209:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 209:
US-10-193-002-209

Query Match 100.0%; Score 96; DB 14; Length 802;
Best Local Similarity 100.0%; Pred. No. 3.7e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFLLPDAQSIQAAAAGFASK 20
|||||
Db 322 NFLLPDAQSIQAAAAGFASK 341

Search completed: March 10, 2004, 12:41:30
Job time : 25.1282 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:57:36 ; Search time 31.6667 seconds
(without alignments)
199.275 Million cell updates/sec

Title: US-10-044-703-47

Perfect score: 96

Sequence: 1 NLLPDAQSIQAAAAGFASK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	96	100.0	374	16 Q7U0Z8	Q7U0Z8 mycobacteri
2	83	86.5	374	2 Q49590	Q49590 mycobacteri
3	78	81.2	374	2 Q49589	Q49589 mycobacteri
4	46	47.9	149	16 Q7WMD5	Q7WMD5 haemophilus
5	46	47.9	955	12 Q68558	Q68558 human cytom
6	45	46.9	281	16 Q8XRZ8	Q8XRZ8 xanthomonas
7	45	46.9	528	16 Q8XRK4	Q8XRK4 ralstonia s
8	44	45.8	319	16 Q911Q6	Q911Q6 pseudomonas
9	44	45.8	353	5 Q8MRN0	Q8MRN0 drosophila
10	44	45.8	410	16 Q89M92	Q89M92 bradyrhizob
11	44	45.8	686	16 Q7UPX4	Q7UPX4 rhodospirill
12	44	45.8	2023	5 Q9V5Z9	Q9V5Z9 drosophila
13	44	45.8	2023	5 Q96542	Q96542 drosophila
14	44	45.8	7257	2 Q9K1Z7	Q9K1Z7 polyangium
15	44	45.8	7257	2 Q9L8C7	Q9L8C7 polyangium
16	43	44.8	231	5 Q8MQM5	Q8MQM5 drosophila

17 43 44.8 241 5 Q9W3X9
18 43 44.8 327 16 Q8Z1T0
19 43 44.8 356 16 Q8P968
20 43 44.8 447 16 Q8XRS3
21 43 44.8 735 13 Q57381
22 43 44.8 911 16 Q88LE1
23 43 44.8 1938 2 Q7WXH0
24 42.5 44.3 218 2 Q8GMB9
25 42 43.8 161 4 Q8NAR1
26 42 43.8 303 17 Q8Z5Z1
27 42 43.8 380 16 Q9HKQ8
28 42 43.8 413 16 Q8CRX0
29 42 43.8 448 16 Q8PD18
30 42 43.8 465 16 Q82M26
31 42 43.8 465 16 Q9L057
32 42 43.8 483 5 Q9N9J4
33 42 43.8 611 16 Q74375
34 42 43.8 698 11 Q8BW69
35 42 43.8 748 11 Q8BZ29
36 41 42.7 157 16 Q8R9N1
37 41 42.7 205 16 Q910X8
38 41 42.7 241 11 Q9Z135
39 41 42.7 252 16 Q7TTT7
40 41 42.7 256 10 Q94CW0
41 41 42.7 281 16 Q97132
42 41 42.7 309 16 Q9HUD8
43 41 42.7 314 16 Q8FSM1
44 41 42.7 332 16 Q99ZX7
45 41 42.7 333 16 Q8K7S4

ALIGNMENTS

RESULT 1
Q7U0Z8
ID Q7U0Z8 PRELIMINARY; PRT; 374 AA.
AC Q7U0Z8;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DE Periplasmic phosphate-binding lipoprotein PSTS1 (PBP-1) (PSTS1).
GN PSTS1 OR MB0959.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12789972;
RA Garnier T., Egliseier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248337; CAD93820.1; -;
KW Complete proteins.
SQ SEQUENCE 374 AA; 38215 MW; 1434968191FP201D CRC64;

Query Match 100.0%; Score 96; DB 16; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLLPDAQSIQAAAAGFASK 20
Db 272 NLLPDAQSIQAAAAGFASK 291

RESULT 2

Q49590

ID Q49590 PRELIMINARY; PRT; 374 AA.

AC Q49590;
 DT 01-NOV-1996 (TREMELrel. 01, Created)
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE PstS subunit of ABC transporter.
 GN PSTS1A.
 OS Mycobacterium intracellulare.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1767;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35761;
 RX MEDLINE=97055782; PubMed=8900068;
 RA Thangaraj H.S., Bull T.J., De Smet K.A.L., Hill M., Rouse D.A.,
 RA Moreno C., Ivanyi J.;
 RT "Duplication of genes encoding the immunodominant 38 kDa antigen in
 RT Mycobacterium intracellulare.";
 RL FEMS Microbiol. Lett. 144:235-240(1996).
 DR EMBL; X95538; CAA64784.1; -;
 DR HSSP; P06128; IIAQ.
 DR GO; GO:0005215; P:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000437; Prok_lipoprot_S.
 DR InterPro; IPR006059; SBP_Bac_1.
 DR Pfam; PF01547; SBP_bac_1; 1.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 SQ SEQUENCE 374 AA; 38088 MW; AF72D5843860B9D7 CRC64;

Query Match 86.5%; Score 83; DB 2; Length 374;
 Best Local Similarity 89.5%; Pred. No. 2.8e-05;
 Matches 17; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 FLPPDAQSIQAAAAGFASK 20
 DB 272 YLLPDAKSIQAAAAGFASK 290
 :||||:|||||
 :||||:|||||

RESULT 3
 Q49589 PRELIMINARY; PRT; 374 AA.
 AC Q49589;
 DT 01-NOV-1996 (TREMELrel. 01, Created)
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE PstS subunit of ABC transporter.
 GN PSTS1A.
 OS Mycobacterium intracellulare.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1767;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35761;
 RX MEDLINE=97055782; PubMed=8900068;
 RA Thangaraj H.S., Bull T.J., De Smet K.A.L., Hill M., Rouse D.A.,
 RA Moreno C., Ivanyi J.;
 RT "Duplication of genes encoding the immunodominant 38 kDa antigen in
 RT Mycobacterium intracellulare.";
 RL FEMS Microbiol. Lett. 144:235-240(1996).
 DR EMBL; X95538; CAA64783.1; -;
 DR HSSP; P06128; IAS4.
 DR GO; GO:0005215; P:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000437; Prok_lipoprot_S.
 DR InterPro; IPR006059; SBP_Bac_1.
 DR Pfam; PF01547; SBP_bac_1; 1.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 SQ SEQUENCE 374 AA; 37998 MW; 45796D4E9F6F513D CRC64;

Query Match 81.2%; Score 78; DB 2; Length 374;
 Best Local Similarity 84.2%; Pred. No. 0.00019;
 Matches 16; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 FLPPDAQSIQAAAAGFASK 20
 DB 273 YLLPNAKSIQAAAAGFASK 191
 :||||:|||||
 :||||:|||||

RESULT 4
 Q7VMD5 PRELIMINARY; PRT; 149 AA.
 AC Q7VMD5;
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE S0S ribosomal protein I9.
 GN RPLI OR HD1048.
 OS Haemophilus ducreyi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=730;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=35000HP / ATCC 700724;
 RA Munson R.S.Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
 RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
 RT "The complete genome sequence of Haemophilus ducreyi";
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AS017153; AAP95922.1; -;
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 149 AA; 15364 MW; 38F6EB72336FAB15 CRC64;

Query Match 47.9%; Score 46; DB 16; Length 149;
 Best Local Similarity 44.4%; Pred. No. 14;
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 NFLLPDAQSIQAAAAGFASK 18
 DB 28 NFLLPQKAVWATAANIA 45
 :||||:|||||
 :||||:|||||

RESULT 5
 Q68558 PRELIMINARY; PRT; 955 AA.
 AC Q68558;
 DT 01-NOV-1996 (TREMELrel. 01, Created)
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE UL105 protein.
 GN UL105.
 OS Human cytomegalovirus (strain Towne).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Cytomegalovirus.
 OX NCBI_TaxID=10363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Towne;
 RA Pari G.S.;
 RT "Characterization of human cytomegalovirus UL105 gene and
 RT identification of its putative protein product.";
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U51564; AAA98670.1; -;
 DR GO; GO:0005524; P:ATP binding; IEA.
 DR GO; GO:0004386; P:helicase activity; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR InterPro; IPR003840; Herpes_helicase.
 DR Pfam; PF02689; Herpes_Helicase; 1.
 SQ SEQUENCE 955 AA; 106539 MW; 6BCE71E24AE36A41 CRC64;

Query Match 47.9%; Score 46; DB 12; Length 955;
 Best Local Similarity 62.5%; Pred. No. 1e+02;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 LPDAQSIQAAAAGFASK 19
 :||||:|||||
 :||||:|||||

```

Db      633 LPDVSSICAAAGHAA 648
RESULT 6
Q8PJ28 PRELIMINARY; PRT; 281 AA.
AC
Q8PJ28;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein XAC2379.
GN XAC2379.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN
SEQUENCE FROM N.A.
RP STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=1204217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AB011874; AAM37231.1;
DR InterPro; IPR000437; Prok_lipoProtein_S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 281 AA; 2955 MW; 0422FA793E08926 CRC64;

Query Match 46.9%; Score 45; DB 16; Length 281;
Best Local Similarity 71.4%; Pred. No. 40;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 DAQSIQAAAAGFAS 19
| | | | |
| | | | |
Db 43 DAQSVAAVAAAGRAS 56
| | | | |
| | | | |

RESULT 7
Q8XRK4 PRELIMINARY; PRT; 528 AA.
AC Q8XRK4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Probable ketoglutarate semialdehyde dehydrogenase protein
DE (EC 1.2.-).
GN RSP0827 OR RS05368.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN
SEQUENCE FROM N.A.
RP STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salancubut M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,

```

```

RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguiet P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646081; CAD17978.1;
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldenhyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR Oxidoreductase; Plasmid; Complete proteome.
SQ SEQUENCE 528 AA; 55211 MW; 380E493B6587A685 CRC64;

Query Match 46.9%; Score 45; DB 16; Length 528;
Best Local Similarity 66.7%; Pred. No. 80;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 FLLPDAOSIQAAAAG 16
| | | | |
| | | | |
Db 272 FLLPDALARRAALG 286
| | | | |
| | | | |

RESULT 8
Q911Q6 PRELIMINARY; PRT; 319 AA.
AC Q911Q6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein PA2211.
GN PA2211.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN
SEQUENCE FROM N.A.
RP STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Miziouchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004647; AAG05599.1;
DR PIR; B83370; B83370.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 319 AA; 36100 MW; 032763ECABF67646 CRC64;

Query Match 45.8%; Score 44; DB 16; Length 319;
Best Local Similarity 64.3%; Pred. No. 67;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 PDAQSIQAAAAGFA 18
| | | | |
| | | | |
Db 105 PDAAEVLAAQAGFA 118
| | | | |
| | | | |

RESULT 9
Q8MRNO PRELIMINARY; PRT; 353 AA.
AC Q8MRNO;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE GH14582p.

```

GN E(PC) OR CG7776.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.,
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY119519; AAM50173.1; -
DR FlyBase; FBgn000581; E(PC).
SQ SEQUENCE 353 AA; 36184 MW; 6A024513A84845C4 CRC64;

Query Match 45.8%; Score 44; DB 5; Length 353;
Best Local Similarity 56.2%; Pred. No. 75;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 NFFLPDAQSIQAAAAG 16
||| : : : : :
Db 107 NFTIATASELQAAAAG 122

RESULT 10
Q89M92 PRELIMINARY; PRT; 410 AA.
ID Q89M92;
AC Q89M92;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE B114301 protein.
GN B114301.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005950; BAC49566.1; -
DR InterPro; IPR002005; NAD_RS.
KW Complete proteome.
SQ SEQUENCE 410 AA; 44842 MW; 4D2C713F8EDC6C47 CRC64;

Query Match 45.8%; Score 44; DB 16; Length 410;
Best Local Similarity 56.2%; Pred. No. 88;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 3 LLPDAQSIQAAAAGFA 18
||| : : : : :
Db 218 VLGDQSLQQTAAAGWS 233

RESULT 11
Q7UPX4 PRELIMINARY; PRT; 686 AA.
ID Q7UPX4
AC Q7UPX4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN RB6668.
OC Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR EMBL; BX294144; CAD74934.1; -
DR Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 686 AA; 76135 MW; E32412ECB378B2A8 CRC64;

Query Match 45.8%; Score 44; DB 16; Length 686;
Best Local Similarity 56.2%; Pred. No. 1.6e+02;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 3 LLPDAQSIQAAAAGFA 18
||| : : : : :
Db 448 MLAEAAQAAAAGLA 463

RESULT 12
Q9V5Z9 PRELIMINARY; PRT; 2023 AA.
ID Q9V5Z9;
AC Q9V5Z9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG7776 protein.
GN E(PC) OR CG7776.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton G.G., Mortman J.K., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siders-Klamas I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*,"
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreria S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houch J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattel B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel V., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
RT "Sequencing of *Drosophila melanogaster* genome,"
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman D., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
RT "Annotation of *Drosophila melanogaster* genome,"
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003826; AAW68743.1; -
DR FlyBase; FBgn0000581; E(PC).
SQ SEQUENCE 2023 AA; 220868 MW; 4D3EB2DB8623550B CRC64;
Query Match 45.8%; Score 44; DB 5; Length 2023;
Best Local Similarity 56.2%; Pred. No. 5.1e+02;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 NFFLPDAQSIQAAAAAG 16
Db 1777 NFTIATASELQAAAAAG 1792
RESULT 13
Q96542 ID Q96542 PRELIMINARY; PRT; 2023 AA.
AC Q96542;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Enhancer of POLYCOMB.
GN E(PC) OR E OR CG7776.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; *Drosophila*.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Imaginal disks;
RX MEDLINE=98407961; PubMed=9735366;
RA Stankunas K., Berger J., Ruse C., Sinclair D.A.R., Randazzo F.,
RA Brock H.W.;
RT "The enhancer of polycomb gene of *Drosophila* encodes a chromatin
RT protein conserved in yeast and mammals,"
RL Development 125:4055-4066(1998).
DR EMBL; AF079764; AAC64271.1; -
DR FIR; T13154; T13154.
DR FlyBase; FBgn0000581; E(PC).
SQ SEQUENCE 2023 AA; 220678 MW; 4A8E71E5A998249A CRC64;
Query Match 45.8%; Score 44; DB 5; Length 2023;
Best Local Similarity 56.2%; Pred. No. 5.1e+02;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 NFFLPDAQSIQAAAAAG 16
Db 1777 NFTIATASELQAAAAAG 1792
RESULT 14
Q9K1Z7 ID Q9K1Z7 PRELIMINARY; PRT; 7257 AA.
AC Q9K1Z7;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Epod.
GN EPD.
OS *Polyangium cellululosum*.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Sorangineae; Polyangiaceae; *Polyangium*.
OX NCBI_TaxID=56;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SMP44;
RX MEDLINE=20293058; PubMed=10831849;
RA Julien B., Shah S., Ziermann R., Goldman R., Katz L., Khosla C.;
RT "Isolation and characterization of the epothilone biosynthetic gene
RT cluster from *Sorangium cellululosum*,"
RL Gene 249:153-160(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SMP44;
RX MEDLINE=20115953; PubMed=10649995;
RA Tang L., Shah S., Chung L., Carney J., Katz L., Khosla C., Julien B.;
RT "Cloning and heterologous expression of the epothilone gene cluster,"
RL Science 287:640-642(2000).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
DR EMBL; AF217189; AAF62883.1; -
DR HSSP; P25715; 1MLA.
DR GO; GO:0004024; P:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0003960; P:NADPH:quinone reductase activity; IEA.
DR GO; GO:0016491; P:oxidoreductase activity; IEA.
DR GO; GO:0004289; P:subtilase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0008270; P:zinc ion binding; IEA.
DR GO; GO:0004314; P:[acyl-carrier protein] S-malonyltransferase.; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001227; AcTrans.
DR InterPro; IPR002085; Acn_zn_family.
DR InterPro; IPR004410; FabD.
DR InterPro; IPR000794; Ketoacyl synth.
DR InterPro; IPR000209; Peptidase_S8.

```

DR InterPro; IPR006162; Ppantne_S.
DR InterPro; IPR006163; Pp bind.
DR InterPro; IPR002364; QOR zeta crystal.
DR Pfam; PF00698; Acyl trans; 4.
DR Pfam; PF00107; ADH zinc N; 2.
DR Pfam; PF00109; ketoacyl-synt; 4.
DR Pfam; PF02801; ketoacyl-synt_C; 4.
DR Pfam; PF00550; pp-binding; 4.
DR TIGFAMS; TIGR00128; fabd; 4.
DR PROSITE; PS00075; ACP DOMAIN; 4.
DR PROSITE; PS00606; B KETOACYL SYNTHASE; 4.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR PROSITE; PS01162; QOR ZETA CRYSTAL; 1.
DR PROSITE; PS00136; SUBTILASE-ASP; 1.
DR Oxidoreductase; Phosphopantetheine; Transferase.
SQ SEQUENCE 7257 AA; 764877 MW; 3A536509FA2FCE99 CRC64;

Query Match 45.8%; Score 44; DB 2; Length 7257;
Best Local Similarity 61.5%; Pred. No. 2.1e+03;
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 8 QSIQAAAAGPASK 20
DB 733 QAVQAIAAGFAAR 745

RESULT 15
Q9L8C7 PRELIMINARY; PRT; 7257 AA.
AC Q9L8C7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 25, Last annotation update)
DE Polyketide synthase.
GN EPOC
OS Polyangium cellulosum.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Sorangineae; Polyangiaceae; Polyangium.
OX NCBI_TaxID=56;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=So ce90;
RX MEDLINE=20130945; PubMed=10662695;
RA Molnar I., Schupp T., Ono M., Zirkle R.E., Milmanow M.,
RA Nowak-Thompson B., Engel N., Toupet C., Stratmann A., Cyr D.D.,
RA Grolach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;
RT "The biosynthetic gene cluster for the microtubule-stabilizing agents
RT epothilones A and B from Sorangium cellulosum So ce90.";
RL Chem. Biol. 7:97-109(2000).
DR EMBL; AF210843; AAF26921.1; -.
DR HSBP; P25715; IMLA.
DR GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0003960; F:NADPH:quinone reductase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0004314; F:[acyl-carrier protein] S-malonyltransferase. .; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001227; Ac trans.
DR InterPro; IPR002085; Adh zn_family.
DR InterPro; IPR004410; Fabd.
DR InterPro; IPR000794; Ketoacyl synth.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR006162; Ppantne_S.
DR InterPro; IPR006163; Pp bind.
DR InterPro; IPR002364; QOR zeta crystal.
DR Pfam; PF00698; Acyl trans; 4.
DR Pfam; PF00107; ADH zinc N; 2.
DR Pfam; PF00109; ketoacyl-synt; 4.
DR Pfam; PF02801; ketoacyl-synt_C; 4.
DR Pfam; PF00550; pp-binding; 4.

DR TIGFAMS; TIGR00128; fabd; 4.
DR PROSITE; PS00075; ACP DOMAIN; 4.
DR PROSITE; PS00606; B KETOACYL SYNTHASE; 4.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR PROSITE; PS01162; QOR ZETA CRYSTAL; 1.
DR PROSITE; PS00136; SUBTILASE-ASP; 1.
DR Oxidoreductase; Phosphopantetheine; Transferase.
SQ SEQUENCE 7257 AA; 764877 MW; 3A536509FA2FCE99 CRC64;

Query Match 45.8%; Score 44; DB 2; Length 7257;
Best Local Similarity 61.5%; Pred. No. 2.1e+03;
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 8 QSIQAAAAGPASK 20
DB 733 QAVQAIAAGFAAR 745

Search completed: March 10, 2004, 12:10:50
Job time : 33.7917 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:51:36 ; Search time 6.15385 Seconds
(without alignments)
169.228 Million cell updates/sec

Title: US-10-044-703-47

Perfect score: 96

Sequence: 1 NFLLPDAQSIQAAAGFASK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	374	1 PST1_MYCTU	P15712 mycobacteri
2	45	46.9	379	1 VAT1_TORCA	P19333 torpedo cal
3	43	44.8	327	1 QOR_SALTY	P40783 salmonella
4	42.5	44.3	704	1 TRFE_BOVIN	Q29443 bos taurus
5	42	43.8	191	1 RL9_AGR75	Q8uqf0 agrobacteri
6	42	43.8	450	1 ECFE_SALTY	Q8z9a4 salmonella
7	42	43.8	450	1 ECFE_SALTY	Q8z9a4 salmonella
8	42	43.8	540	1 CH11_RHOL	P29026 rhizopus ol
9	42	43.8	923	1 PRGF_MOUSE	Q00175 mus musculu
10	42	43.8	923	1 PRGF_MOUSE	Q00175 mus musculu
11	41	42.7	149	1 RL9_XYLF	Q63449 rattus norv
12	41	42.7	707	1 DCOR_LEIDO	Q87a85 xyliella fas
13	41	42.7	986	1 BMPI_HUMAN	P27116 leishmania
14	41	42.7	991	1 BMPI_MOUSE	P13497 homo sapien
15	40.5	42.2	580	1 MM14_PIG	P98063 mus musculu
16	40	41.7	149	1 RL9_VIRCH	Q9xt90 sus scrofa
17	40	41.7	149	1 RL9_XANAC	Q9kuy9 vibrio chol
18	40	41.7	149	1 RL9_XANAC	Q9kuy9 vibrio chol
19	40	41.7	816	1 HDA7_CAEEL	Q8pac0 xanthomonas
20	40	41.7	1022	1 TLD_BRARE	O17323 caenorhabdi
21	40	41.7	1050	1 ITA5_XENLA	O56274 xenopus lae
22	40	41.7	1473	1 OVOS_CHICK	Q62740 gallus gall
23	39.5	41.1	518	1 PVR3_SYNEL	Q8dtn5 s bifunctio
24	39.5	41.1	582	1 NM14_HUMAN	P50281 homo sapient
25	39	40.6	143	1 VAT1_MOUSE	Q62465 mus musculu
26	39	40.6	143	1 RL9_XYLF	Q9paf9 xyliella fas
27	39	40.6	174	1 Y433_ASRPE	Q9yf05 aeropyrum p
28	39	40.6	194	1 RL9_CAUCR	Q9a7q3 caulobacter
29	39	40.6	222	1 RL9_RHOCA	O68128 rhodobacter
30	39	40.6	247	1 HIST_RHIL	Q98ct3 rhizobium l
31	39	40.6	300	1 VAT1_HUMAN	Q99336 homo sapien
32	39	40.6	434	1 YU18_MYCTU	P31500 mycobacteri
33	39	40.6	435	1 YU21_MYCTU	O53268 mycobacteri

34 39 40.6 671 1 AMO1_ASPNG Q12556 aspergillus
35 39 40.6 721 1 BBS2_RAT Q99mh9 rattus norv
36 39 40.6 956 1 HEL1_HCVWA P16736 human cytom
37 39 40.6 2514 1 POLN_ONKVG P13888 o'nyong-nyo
38 38.5 40.1 193 1 CLPP_FUSNN Q8rhj8 fuscobacteri
39 38.5 40.1 218 1 UVRY_ECOLD Q8xbd4 escherichia
40 38.5 40.1 218 1 UVRY_ECOLD P07027 escherichia
41 38.5 40.1 244 1 EXBB_SALTY P18950 salmonella
42 38.5 40.1 261 1 COBK_SYNY3 P72711 synechocyst
43 38 39.6 94 1 YAI1_ECOLD P36768 escherichia
44 38 39.6 128 1 CYCP_RHOFU P00151 rhodospiril
45 38 39.6 128 1 CYCP_RHOMO P00152 rhodospiril

ALIGNMENTS

RESULT 1
ID PST1_MYCTU STANDARD; PRT; 374 AA.
AC P15712; O05868;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phosphate-binding protein 1 precursor (PBP-1) (Protein
DE antigen B) (PAB) (Antigen Ag78)
GN PST1 OR PHO1 OR RV0934 OR WT0961 OR MTCV08D9.05C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8307568; PubMed=2545626;
RA Andersen A.B., Hansen E.B.;
RT "Structure and mapping of antigenic domains of protein antigen b, a
RT 38,000-molecular-weight protein of Mycobacterium tuberculosis.";
RL Infect. Immun. 57:2481-2488(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=2206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Petersen J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Emolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
RN [4]
RP FUNCTION: Required for binding-protein-mediated phosphate
RP transport.
RN [5]
RP SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
RP (Probable).
RN [6]
RP SIMILARITY: Belongs to the pbs family.
RN [7]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN [8]
RP between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

CC -----
 CC EMBL; M30046; AA25374.1; --
 CC EMBL; 295209; CA08484.1; --
 CC EMBL; AE006981; AAK45208.1; --
 CC DR PIR; F70584; P70584.
 CC DR HSSP; P06128; 11XH.
 CC DR TIGR; MT0961; --
 CC DR TubercuList; Rv0934; --
 CC DR InterPro; IPR000437; Prok lipoprot_S.
 CC DR InterPro; IPR006059; SBP_Bac_1.
 CC DR Pfam; PF01547; SBP_Bac_1; 1.
 CC DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 CC KW Phosphate transport; Transport; Membrane; Lipoprotein; Signal;
 CC Antigen; Complete proteome; Palmitate.
 CC FT SIGNAL 1 23 POTENTIAL.
 CC FT CHAIN 24 374 PHOSPHATE-BINDING PROTEIN 1.
 CC FT LIPID 24 24 N-palmitoyl cysteine (Potential).
 CC FT LIPID 24 24 S-diacylglycerol cysteine (Potential).
 CC SQ SEQUENCE 374 AA; 38243 MW; 6334968191FF38AA CRC64;

Query Match 100.0%; Score 96; DB 1; Length 374;
 Best Local Similarity 100.0%; Pred. No. 1.7e-08; Mismatches 0; Indels 0; Gaps 0;
 Matches 20; Conservative 0;

QY 1 NLLPDAQSIQAAAGFASK 20
 DB 272 NLLPDAQSIQAAAGFASK 291

RESULT-2

VAT1_TORCA
 ID VAT1_TORCA STANDARD; PRT; 379 AA.
 AC P1933;
 RC TISSUE-Electric lobe;
 RX MEDLINE=9016593; PubMed=2483112;
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Synaptic vesicle membrane protein VAT-1.
 OS Torpedo californica (Pacific electric ray).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 CC Elasmobranchii; Squales; Hymnosquales; Pristiogastera; Batoidae;
 CC Torpediniformes; Torpedinidae; Torpedinidae; Torpedo.
 CC NCBI_TaxID=7787;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-Electric lobe;
 RX MEDLINE=9016593; PubMed=2483112;
 RA Linial M., Miller K., Scheller R.H.;
 RT "VAT-1: an abundant membrane protein from Torpedo cholinergic
 RT synaptic vesicles.";
 RL Neuron 2:1265-1273 (1989).
 CC -!- FUNCTION: May play a central role in the functions mediated by
 CC specific classes of synaptic vesicles.
 CC -!- SUBCELLULAR LOCATION: Membrane.
 CC -!- TISSUE SPECIFICITY: Cholinergic synaptic vesicles.
 CC -!- MISCELLANEOUS: Synthesized in the neuronal cell bodies and
 CC transported to the terminals. Each vesicle contains approximately
 CC 28 molecules of VAT-1.
 CC -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY. QUINONE OXIDOREDUCTASE SUBFAMILY. STRONG, TO MAMMALIAN
 CC VAT-1 HOMOLOGS.
 CC PIR; JN0013; JN0013.
 CC InterPro; IPR002085; Adh_zn_family.
 CC InterPro; IPR002364; QOR_zeta_crystal.
 CC Pfam; PF00107; ADH_zinc_N; 1.
 CC PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
 CC OXidoreductase; Zinc; Synapse; Membrane; Phosphorylation.
 CC MOD_RS 273 273 PHOSPHORYLATION (POTENTIAL).
 CC SEQUENCE 379 AA; 41593 MW; FA4ADAI7E657F09C CRC64;

Query Match 46.9%; Score 45; DB 1; Length 379;
 Best Local Similarity 64.3%; Pred. No. 6.2; Mismatches 1; Indels 0; Gaps 0;
 Matches 9; Conservative 1;

QY 2 FLPPDAQSIQAAAA 15
 DB 130 FLPPDGMGFQAAAA 143

RESULT 3

QOR_SALTY
 ID QOR_SALTY STANDARD; PRT; 327 AA.
 AC F40783;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase) (zeta-
 DE crystallin homolog protein).
 CC QOR OR STM4245.
 GN Salmonella typhimurium.
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 CC NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856 (2001).
 RN [2]
 RP SEQUENCE OF 1-168 FROM N.A.
 RX MEDLINE=88227847; PubMed=2836367;
 RA Wong A., Kean L., Maurer R.;
 RT "Sequence of the dnaB gene of Salmonella typhimurium.";
 RL J. Bacteriol. 170:2668-2675 (1988).
 CC -!- CATALYTIC ACTIVITY: NADPH + quinone = NADP(+) + semiquinone.
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase
 CC family. Quinone oxidoreductase subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

CC EMBL; AE008898; AAL23069.1; --
 CC EMBL; J03390; -; NOT ANNOTATED_CDS.
 CC HSSP; P28304; 1QOR.
 CC StyGene; SG10511; gor.
 CC InterPro; IPR002085; Adh_zn_family.
 CC InterPro; IPR002364; QOR_zeta_crystal.
 CC Pfam; PF00107; ADH_zinc_N; 1.
 CC PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
 CC OXidoreductase; NADP; Zinc; Complete proteome.
 CC FT CONFLICT 92 92 L -> V (IN REF. 2).
 CC FT CONFLICT 148 148 A -> R (IN REF. 2).
 CC SQ SEQUENCE 327 AA; 35151 MW; 602F4CB6E6A805F1 CRC64;

Query Match 44.8%; Score 43; DB 1; Length 327;
 Best Local Similarity 60.0%; Pred. No. 12; Mismatches 2; Indels 4; Gaps 0;
 Matches 9; Conservative 2;

QY 3 LLPDAQSIQAAAGF 17

RA Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wolman C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursun J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RL Agrobacterium tumefaciens C58";
RL Science 294:2323-2328(2001).
CC -!- FUNCTION: Binds to the 23S rRNA (By similarity).
CC -!- SIMILARITY: Belongs to the L9 family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB009072; AAL42101.1; -;
CC EMBL; AB008039; AAK86897.1; -;
CC PIR; AG2710; AG2710.
CC PIR; H97492; H97492.
CC HMAP; MF 00503; 1;
CC InterPro; IPR009027; L9_N_like.
CC InterPro; IPR000244; Ribosomal_L9.
CC Pfam; PF03948; Ribosomal_L9_C; 1.
CC Pfam; PF01281; Ribosomal_L9_N; 1.
CC DR TIGRFAMs; TIGR00158; L9; 1.
CC DR PROSITE; PS00651; RIBOSOMAL_L9; 1.
CC Ribosomal protein; rRNA-binding; Complete proteome.
CC KW Ribosomal protein; rRNA-binding; Complete proteome.
CC SQ SEQUENCE 191 AA; 20922 MW; 31355D491FC846BF CRC64;

Query Match 43.8%; Score 42; DB 1; Length 191;
Best Local Similarity 53.3%; Pred. No. 10;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLLPDAQSIQAAAA 15
DQ 28 NLLPQGGKALANAA 42

RESULT 6
ID ECPE SALT STANDARD; PRT; 450 AA.
AC Q823A4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protease ecfe (EC 3.4.24.-).
GN ECPE OR STY0246 OR T0224.
OS Salmonella typhi
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.B., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moulé S., O'Gara P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RL enterica serovar Typhi CT18";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;

RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: Degrades both heat shock sigma factors rpoE and rpoH (By
CC similarity).
CC -!- COFACTOR: Zinc (Probable).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family M50B.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL627266; CAD98681.1; -;
CC EMBL; AB016834; AAO67954.1; -;
CC MEROPS; M50.004; -;
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR004387; Pept_M50_Zn.
CC InterPro; IPR006025; Pept_M_Zn_BS.
CC InterPro; IPR008915; Peptidase_M50.
CC Pfam; PF00595; PDZ; 1.
CC Pfam; PF02163; Peptidase_M50; 1.
CC SMART; SMO0228; PDZ; 2.
CC TIGRFAMs; TIGR00054; TIGR00054; 1.
CC DR PROSITE; PS0106; PDZ; 1.
CC DR PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Zinc; Transmembrane; Inner membrane;
CC KW Complete proteome.
CC FT METAL 22 22 ZINC (CATALYTIC) (POTENTIAL).
CC FT ACT SITE 23 23 POTENTIAL.
CC FT METAL 26 26 ZINC (CATALYTIC) (POTENTIAL).
CC FT TRANSMEM 98 120 POTENTIAL.
CC FT TRANSMEM 376 398 POTENTIAL.
CC FT TRANSMEM 426 445 POTENTIAL.
CC FT DOMAIN 199 291 PDZ.
CC SQ SEQUENCE 450 AA; 49211 MW; BD05018CE5058153 CRC64;

Query Match 43.8%; Score 42; DB 1; Length 450;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 LLPDAQSIQAAAA 18
DQ 287 LTPDTKSVNGKAEFGA 302

RESULT 7
ID ECPE SALT STANDARD; PRT; 450 AA.
AC Q82RF1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protease ecfe (EC 3.4.24.-).
GN ECPE OR STM0223.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 L22.";
 RL Nature 413:852-856(2001).
 CC -!- FUNCTION: Degrades both heat shock sigma factors rpoE and rpoH (By
 CC similarity).
 CC -!- COFACTOR: Zinc (Probable).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (By similarity).
 CC -!- SIMILARITY: Belongs to peptidase family M50B.
 CC -!- SIMILARITY: Contains 1 PDZ/DRH domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE008705; AAL19187.1; -.
 CC MEROPS; M50.004; -.
 CC StyGene; SG22222; ecFE.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR004387; Pept_M50_Zn.
 DR InterPro; IPR006025; Pept_M_Zn_BS.
 DR InterPro; IPR008915; Peptidase_M50.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF02163; Peptidase_M50; 1.
 DR SMART; SMO0228; PDZ; 2.
 DR TIGRFAMs; TIGR00054; TIGR00054; 1.
 DR PROSITE; PS50106; PDZ; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Hydrolyase; Metalloprotease; Zinc; Transmembrane; Inner membrane;
 KW Complete proteome.
 FT METAL 22 22 ZINC (CATALYTIC) (POTENTIAL).
 FT ACT SITE 23 23 POTENTIAL.
 FT METAL 26 26 ZINC (CATALYTIC) (POTENTIAL).
 FT TRANSMEM 98 120 POTENTIAL.
 FT TRANSMEM 376 398 POTENTIAL.
 FT TRANSMEM 426 445 POTENTIAL.
 FT DOMAIN 199 291 PDZ.
 SQ SEQUENCE 450 AA; 49184 MW; 50A748A774A70B7B CRC64;
 Query Match 43.8%; Score 42; DB 1; Length 450;
 Best Local Similarity 50.0%; Pred. No. 24;
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 QY 3 LLEPDAQSIQAAAAAGFA 18
 DB 287 LTPDTKSVNGKAGFA 302
 RESULT 8
 CHIL_RHIOL
 ID CHIL_RHIOL STANDARD; PRT; 540 AA.
 AC P29026;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE Chitinase 1 precursor (EC 3.2.1.14).
 GN CHIL.
 OS Rhizopus oligosporus.
 OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
 OC Rhizopus.
 OC NCBI_TaxID=4847;
 [1]
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 23-52.
 RC STRAIN=Saigo IFO 8631;
 RX MEDLINE=93054356; PubMed=1429462;
 RA Yanai K., Takaya N., Kojima N., Horiuchi H., Ohta A., Takagi M.;

RT "Purification of two chitinases from *Rhizopus oligosporus* and
 RT isolation and sequencing of the encoding genes.";
 RL J. Bacteriol. 174:7398-7406(1992).
 CC -!- FUNCTION: Probably involved in the apical growth and branching
 CC of fungal hyphae.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
 CC acetyl-D-glucosamine polymers of chitin.
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Secreted (Probable).
 CC -!- PTM: O-glycosylated.
 CC -!- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl
 CC hydrolases).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D10157; BAA01021.1; -.
 CC PIR; A47022; A47022.
 DR HSSP; P23472; 2HVM.
 DR InterPro; IPR001223; Glyco_hydro_18.
 DR InterPro; IPR001579; Glyco_hydro_18AS.
 DR InterPro; IPR005089; Glyco_hydro_18C.
 DR Pfam; PF03427; CBM_19; 1.
 DR Pfam; PF00704; Glyco_hydro_18; 1.
 DR PROSITE; PS01095; CHITINASE_18; 1.
 KW Hydrolyase; Glycosidase; Chitin degradation; Signal; Zymogen;
 KW Glycoprotein; Chitin-binding.
 FT SIGNAL 1 22
 FT CHAIN 23 417 CHITINASE 1.
 FT PROPEP 418 540 POTENTIAL.
 FT DOMAIN 23 312 CATALYTIC.
 FT DOMAIN 313 352 SER/THR-RICH.
 FT DOMAIN 353 404 CHITIN-BINDING, HIGH AFFINITY.
 FT DOMAIN 405 417 C-TERMINAL VARIABLE REGION.
 FT ACT SITE 166 166 PROTON DONOR (BY SIMILARITY).
 SQ SEQUENCE 540 AA; 56226 MW; C7A8590D03881F1B CRC64;
 Query Match 43.8%; Score 42; DB 1; Length 540;
 Best Local Similarity 50.0%; Pred. No. 28;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 5 PDAQSIQAAAAAGFASK 20
 DB 415 PSASALEIAAASFGSK 430
 RESULT 9
 PRGR_MOUSE
 ID PRGR_MOUSE STANDARD; PRT; 923 AA.
 AC Q00175;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Progesterone receptor (PR).
 GN PR OR NR3C3 OR PR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=91299759; PubMed=2069958;
 RA Schott D.R., Shyamala G., Schneider W., Parry G.;
 RT "Molecular cloning, sequence analyses, and expression of
 RT complementary DNA encoding murine progesterone receptor.";
 RL Biochemistry 30:7014-7020(1991).
 [2]
 RN SEQUENCE OF 1-9 FROM N.A.

```

RC STRAIN=129/SV;
RA MEDLINE=95100931; PubMed=7802637;
RX Hagihara K., Wu-Peng X.S., Funabashi T., Kato J., Pfaff D.W.;
RT "Nucleic acid sequence and DNase hypersensitive sites of the 5'
RL Biochem. Biophys. Res. Commun. 205:1093-1101(1994).
CC -!- FUNCTION: The steroid hormones and their receptors are involved in
CC the regulation of eukaryotic gene expression and affect cellular
CC proliferation and differentiation in target tissues.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DOMAIN: Composed of three domains: a modulating N-terminal domain,
CC a DNA-binding domain and a C-terminal steroid-binding domain.
CC -!- SIMILARITY: Belongs to the nuclear hormone receptor family. NR3
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M68915; AAA39971.1; -.
CC EMBL; U12644; AAA66067.1; -.
CC PIR; A39596; A39596.
CC HSSP; P06401; 1A28.
CC TRANSFAC; T04680; -.
CC MGD; MGI:197567; Pgr.
CC InterPro; IPR000536; Hormone_rec_lig.
CC InterPro; IPR000128; Progesterone_receptor.
CC InterPro; IPR001723; Steroid_receptor.
CC InterPro; IPR008946; Str_ncl_receptor.
CC InterPro; IPR001628; Znf_C4steroid.
CC Pfam; PF00104; hormone_rec; 1.
CC Pfam; PF02181; Prog_receptor; 1.
CC Pfam; PF00105; zf-C4; 1.
CC PRINTS; PR00398; STRDHOMONER.
CC PRINTS; PR00047; STROIDFINGER.
CC ProDom; PD000035; Znf_C4steroid; 1.
CC SMART; SM00430; HOL1; 1.
CC SMART; SM00399; Znf_C4; 1.
CC PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
CC Receptor; Transcription regulation; DNA-binding; Nuclear protein;
CC Zinc-finger; Steroid-binding.
CC DOMAIN 1 556 MODULATING, PRO-RICH.
CC FT DNA BIND 557 622 NUCLEAR RECEPTOR-TYPE.
CC FT ZN_FING 557 577 C4-TYPE.
CC FT ZN_FING 593 617 C4-TYPE.
CC FT DOMAIN 671 923 STEROID-BINDING.
CC FT DOMAIN 184 188 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC SQ SEQUENCE 923 AA; 99073 MW; 9415F1ED343BEE3F CRC64;

Query Match 43.8%; Score 42; DB 1; Length 923;
Best Local Similarity 42.1%; Pred. No. 48;
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 1 NFFLPDAQSQIAAAGFAS 19
Db 533 NYLRPDSEASQSQYGFDS 551

RESULT 10
PRGR_RAT
ID PRGR_RAT STANDARD; PRT; 923 AA.
AC Q63449;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Progesterone receptor (PR).
GN PGR OR NR3C3.
OS Rattus norvegicus (Rat).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Sprague-Dawley; TISSUE=Placenta;
RA MEDLINE=94130817; PubMed=8299566;
RA Park-Sarge O.K., Mayo K.E.;
RT "Regulation of the progesterone receptor gene by gonadotropins and
RL Endocrinology 134:709-718(1994).
CC -!- FUNCTION: The steroid hormones and their receptors are involved in
CC the regulation of eukaryotic gene expression and affect cellular
CC proliferation and differentiation in target tissues.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DOMAIN: Composed of three domains: a modulating N-terminal domain,
CC a DNA-binding domain and a C-terminal steroid-binding domain.
CC -!- SIMILARITY: Belongs to the nuclear hormone receptor family. NR3
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; L16922; AAA19916.1; -.
CC PIR; I53280; I53280.
CC HSSP; P06401; 1A28.
CC TRANSFAC; T04681; -.
CC InterPro; IPR000536; Hormone_rec_lig.
CC InterPro; IPR000128; Progesterone_receptor.
CC InterPro; IPR001723; Steroid_receptor.
CC InterPro; IPR008946; Str_ncl_receptor.
CC InterPro; IPR001628; Znf_C4steroid.
CC Pfam; PF00104; hormone_rec; 1.
CC Pfam; PF02161; Prog_receptor; 1.
CC Pfam; PF00105; zf-C4; 1.
CC PRINTS; PR00398; STRDHOMONER.
CC PRINTS; PR00047; STROIDFINGER.
CC ProDom; PD000035; Znf_C4steroid; 1.
CC SMART; SM00430; HOL1; 1.
CC SMART; SM00399; Znf_C4; 1.
CC PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
CC Receptor; Transcription regulation; DNA-binding; Nuclear protein;
CC Zinc-finger; Steroid-binding.
CC DOMAIN 1 556 MODULATING, PRO-RICH.
CC FT DNA BIND 557 622 NUCLEAR RECEPTOR-TYPE.
CC FT ZN_FING 557 577 C4-TYPE.
CC FT ZN_FING 593 617 C4-TYPE.
CC FT DOMAIN 671 923 STEROID-BINDING.
CC FT DOMAIN 184 188 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC SQ SEQUENCE 923 AA; 99407 MW; 05384B9656BF22DC CRC64;

Query Match 43.8%; Score 42; DB 1; Length 923;
Best Local Similarity 42.1%; Pred. No. 48;
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 1 NFFLPDAQSQIAAAGFAS 19
Db 533 NYLRPDSEASQSQYGFDS 551

RESULT 11
RL9_XYLF
ID RL9_XYLF STANDARD; PRT; 149 AA.
AC Q87A85;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 50S ribosomal protein L9.
GN RPL1 OR PDI943.

```


FT VARSPLIC 623 986 Missing (in isoform BMP1-5).
 FT FTID=VSP_005466.
 FT DKECKNDGCGQD -> GGELFGLLGHPPRRP (in
 FT isoform BMP1-6).
 FT FTID=VSP_005467.
 FT Missing (in isoform BMP1-6).
 FT FTID=VSP_005468.
 FT DKECKNDGCGQDVCNTFTSGVCCRCRSGFLVHNDKDC
 FT EAGCDHKVSTSTITSPNPDYKPKCKCTWAIISTPGHR
 FT VGLTFMEDISSPECAVDHLEVDGDKAKVLFGRFCG
 FT -> VLEAGDRHSHLGLLELCPHALVDTVPAPSGALHGD
 FT THATHTHVHTCHPIAQETCRGPGLGASRLSPQGRCHLTLA
 FT PQSGSLVDWDTHRGDKPKRRRSKLTSLTPAFRGIWA
 FT L (in isoform BMP1-7).
 FT FTID=VSP_005469.
 FT Missing (in isoform BMP1-7).
 FT FTID=VSP_005470.
 FT D -> N (IN REF. 4).
 FT CONFLICT 748 748 R -> S (IN REF. 4).
 FT CONFLICT 934 934 R -> S (IN REF. 4).
 FT SEQUENCE 986 AA; 111248 MW; F89201913AC3CBEA CRC64;
 SQ

Query Match 42.7%; Score 41; DB 1; Length 986;
 Best Local Similarity 44.4%; Pred. No. 75;
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 LLPDAQSIQAAAGFASK 20
 DB 580 LAPDKRCEACGGFLTK 597

RESULT 14
 BMP1_MOUSE STANDARD; PRT; 991 AA.
 AC P98063;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Bone morphogenetic protein 1 precursor (EC 3.4.24.19) (BMP-1)
 DE (Procollagen C-proteinase) (PCP) (Mammalian tolloid protein) (mrlid).
 GN BMP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Embryo;
 RX MEDLINE=94229342; PubMed=8174772;
 RA Fukagawa M., Noboru S., Hogan B.L.M., Jones C.M.;
 RT "Embryonic expression of mouse bone morphogenetic protein-1 (BMP-1),
 RT which is related to the Drosophila dorsoventral gene tolloid and
 RT encodes a putative astacin metalloendopeptidase.";
 RL Dev. Biol. 163:175-183(1994).
 CC -I- FUNCTION: Cleaves the C-terminal propeptides of procollagen I, II
 CC and III. Induces cartilage and bone formation.
 CC -I- CATALYTIC ACTIVITY: Cleavage of the C-terminal propeptide at
 CC Ala-I-Asp in type I and II procollagens and at Arg-I-Asp in type
 CC III.
 CC -I- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -I- ENZYME REGULATION: Activity is increased by the procollagen C-
 CC endopeptidase enhancer protein.
 CC -I- TISSUE SPECIFICITY: At high levels in embryonic maternal deciduum
 CC and floor plate region of the neural tube. Less in developing
 CC membranous and endochondral bone, submucosa of intestine, dermis
 CC of skin and the mesenchyme of spleen and lung.
 CC -I- SIMILARITY: Belongs to peptidase family M12A.
 CC -I- SIMILARITY: Contains 2 EGF-like domains.
 CC -I- SIMILARITY: Contains 5 CUB domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L24755; AAA37306.1; -.
 CC PIR; I49540; I49540.
 CC HSP; P00736; IAPQ.
 CC MEROPS; M12.005; -.
 CC MGD; MGI:88176; Bm1.
 CC InterPro; IPR000152; Asx_hydroxyl_s.
 CC InterPro; IPR000859; CUB.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR006025; Pept_M_Zn_BS.
 CC InterPro; IPR006026; Peptidase_M.
 CC InterPro; IPR001506; Peptidase_M12A.
 CC Pfam; PF01400; Astacin; 1.
 CC Pfam; PF00431; CUB; 5.
 CC Pfam; PF00008; EGF; 2.
 CC PRINTS; PR00480; ASTACIN.
 CC SMART; SM00042; CUB; 5.
 CC SMART; SM00179; EGF_CA; 2.
 CC SMART; SM00235; ZnMC; 1.
 CC PROSITE; PS00010; ASX_HYDROXYL; 2.
 CC PROSITE; PS01180; CUB; 5.
 CC PROSITE; PS00022; EGF_1; FALSE_NEG.
 CC PROSITE; PS01186; EGF_2; 2.
 CC PROSITE; PS00026; EGF_3; 2.
 CC PROSITE; PS01187; EGF_CA; 2.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC Growth factor; Cytokine; Repeat; Osteogenesis; Chondrogenesis;
 CC KW Hydrolase; Metalloprotease; EGF-like domain; Zinc; Calcium; Signal;
 CC KW Glycoprotein; Zymogen.
 CC SIGNAL 1 25
 CC PROPEP 26 125
 CC CHAIN 126 991
 CC DOMAIN 126 326
 CC FT DOMAIN 327 439
 CC FT DOMAIN 440 551
 CC FT DOMAIN 552 593
 CC FT DOMAIN 596 707
 CC FT DOMAIN 708 748
 CC FT DOMAIN 752 864
 CC FT DOMAIN 865 981
 CC FT METAL 218 218
 CC FT ACT SITE 219 219
 CC FT METAL 222 222
 CC FT METAL 228 228
 CC FT DISULFID 188 191
 CC FT DISULFID 327 353
 CC FT DISULFID 380 402
 CC FT DISULFID 440 466
 CC FT DISULFID 493 515
 CC FT DISULFID 556 568
 CC FT DISULFID 564 577
 CC FT DISULFID 579 592
 CC FT DISULFID 596 622
 CC FT DISULFID 649 671
 CC FT DISULFID 712 723
 CC FT DISULFID 719 732
 CC FT DISULFID 734 747
 CC FT CARBOHYD 96 96
 CC FT CARBOHYD 147 147
 CC FT CARBOHYD 337 337
 CC FT CARBOHYD 368 368
 CC FT CARBOHYD 604 604
 CC SEQUENCE 991 AA; 111607 MW; 68A1847783A0BB9E CRC64;
 SQ

Query Match 42.7%; Score 41; DB 1; Length 991;
 Best Local Similarity 44.4%; Pred. No. 76;
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 LLPDAQSIQAAAGFASK 20

```

DB      585 LAPDRRCCEACGGFLTK 602
RESULT 15
NM14_PIG
ID      QM14_PIG      STANDARD;      PRT;      580 AA.
AC      Q9XIT90;
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DE      Matrix metalloproteinase-14 precursor (EC 3.4.24.80) (MMP-14)
DE      (Membrane-type matrix metalloproteinase 1) (MT-MMP 1) (MTMMP1)
DE      (Membrane-type-1 matrix metalloproteinase) (MT1-MMP) (MTMMP1).
GN      MMP14.
OS      Sus scrofa (Pig).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX      NCBI_TaxID=9823;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=9905929; PubMed=9881602;
RA      Caron C.; Xue J.; Bartlett J.D.;
RT      "Expression and localization of membrane type 1 matrix
RT      metalloproteinase in tooth tissues.";
RL      Matrix Biol. 17:501-511(1998).
CC      -!- FUNCTION: Seems to specifically activate progelatinase A. May thus
CC      trigger invasion by tumor cells by activating progelatinase A on
CC      the tumor cell surface (By similarity). May play a role in the
CC      biomineralization of enamel and dentin.
CC      -!- CATALYTIC ACTIVITY: Endopeptidase activity. Activates
CC      progelatinase A by cleavage of the propeptide at 37-Asn-|-Leu-38.
CC      Other bonds hydrolyzed include 35-Gly-|-Ile-36 in the propeptide
CC      of collagenase 3, and 341-Asp-|-Phe-342, 441-Asp-|-Leu-442 and
CC      354-Gln-|-Thr-355 in the aggrecan interglobular domain.
CC      -!- COFACTOR: Binds 1 zinc ion per subunit, calcium (By similarity).
CC      -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC      -!- TISSUE SPECIFICITY: Highly expressed in developing tooth tissues.
CC      -!- SIMILARITY: Belongs to peptidase family M10A.
CC      -!- SIMILARITY: Contains 1 hemopexin-like domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF067419; AAC38324.1; -.
DR      HSSP; P08254; IHFS.
DR      MEROPS; M10.014; -.
DR      InterPro; IPR000585; Hemopexin.
DR      InterPro; IPR01818; Pept_M10A_M12B.
DR      InterPro; IPR006025; Pept_M_2n_BS.
DR      InterPro; IPR006026; Peptidase_M.
DR      Pfam; PF00045; hemopexin; 4.
DR      Pfam; PF00413; Peptidase_M10; 1.
DR      Pfam; PF03933; Peptidase_M10_N; 1.
DR      PRINTS; PR00138; MATRIXIN.
DR      SMART; SM00120; HX; 4.
DR      SMART; SM00235; ZnMc; 1.
DR      PROSITE; PS00024; HEMOPEXIN; 1.
DR      PROSITE; PS00142; ZINC_PROTEASE; 1.
DR      PROSITE; PS00546; CYSTEINE_SWITCH; 1.
KW      Hydrolase; Metalloprotease; Zinc; Zymogen; Calcium; Signal;
KW      Transmembrane.
FT      SIGNAL          1      28      POTENTIAL.
FT      PROPEP          29      109      ACTIVATION PEPTIDE.
FT      CHAIN           110      580      MATRIX METALLOPROTEINASE-14.
FT      DOMAIN          110      539      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM        540      560      POTENTIAL.
FT      DOMAIN          561      580      CYTOPLASMIC (POTENTIAL).
FT      DOMAIN          314      509
FT      SITE            91
FT      METAL           237
FT      ACT_SITE       238
FT      METAL           241
FT      METAL           247
FT      DISULFID        317
SQ      SEQUENCE 580 AA; 65934 MW; B7B2C2C569A96CAC CRC64;
Query Match          42.2%; Score 40.5; DB 1; Length 580;
Best Local Similarity 36.0%; Pred. No. 54;
Matches 9; Conservative 5; Mismatches 6; Indels 5; Gaps 1;
QY      1 NFLLPD-----AQSIQAAAGFASK 20
      ||||| : : : : :
Db      266 NFVLPPDDRRGIQQLYGSSEGFPTK 290
      ||||| : : : : :
Search completed: March 10, 2004, 12:06:29
Job time : 7.15385 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:58:01 ; Search time 10.7692 Seconds
(without alignments)
178.641 Million cell updates/sec

Title: US-10-044-703-47

Perfect score: 96

Sequence: 1 NFFLPDAQSIQAAAGFASK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96	100.0	374	1 F70584	phosphate specific
2	45	46.9	379	2 JN0013	synaptic vesicle m
3	44	45.8	319	2 B83370	conserved hypothet
4	44	45.8	2023	2 T13154	polycarb protein e
5	43	44.8	327	2 AD1016	NADPH2:quinone red
6	42	43.8	131	2 AG2710	cultivar specific
7	42	43.8	191	2 H97492	hypothetical prote
8	42	43.8	303	2 H84216	probable membrane
9	42	43.8	450	2 A10529	chitinase (EC 3.2.
10	42	43.8	540	2 A47022	hypothetical prote
11	42	43.8	611	2 S76211	progestrone recep
12	42	43.8	923	2 A39596	progestrone recep
13	42	43.8	923	2 I53280	hypothetical prote
14	41	42.7	205	2 F83333	hypothetical prote
15	41	42.7	281	2 H97124	probable short cha
16	41	42.7	309	2 C83017	quinone oxidoreduc
17	41	42.7	337	2 A82309	conserved hypothet
18	41	42.7	375	2 G89972	ornithine decarbox
19	41	42.7	707	2 A42322	procollagen C-endo
20	41	42.7	730	1 BM01	procollagen C-endo
21	41	42.7	823	1 A58788	procollagen C-endo
22	41	42.7	986	1 B58788	procollagen C-endo
23	41	42.7	991	2 I49540	procollagen C-endo
24	41	42.7	1222	2 T22490	hypothetical prote
25	40.5	42.2	760	2 F75530	ribonuclease - Dai
26	40	41.7	149	2 C82333	ribosomal protein
27	40	41.7	288	2 C69869	spore cortex-lytic
28	40	41.7	220	2 T40806	hypothetical prote
29	40	41.7	240	2 F83592	conserved hypothet

hypothetical prote
hypothetical prote
ovostatin precursor
matrix metalloprot
DNA-binding protei
50S ribosomal prot
hypothetical prote
ribosomal protein
probable lpqQ prot
probable lpqQ prot
ribosomal protein
hypothetical prote
hypothetical prote
hypothetical prote
unknown protein en
hypothetical prote

ALIGNMENTS

RESULT 1

F70584
phosphate specific transporter S precursor - Mycobacterium tuberculosis (strain H37RV)
N:Alternate names: antigen b
C:Species: Mycobacterium tuberculosis
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: F70584; JCS103; A42930; A49721; A45820
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Scates, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70584
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-374 <COL>
A:Cross-references: GB:295209; GB:AL123456; NID:G3261750; PIDN:CAB08484.1; PID:G2078049
A:Experimental source: strain H37RV
R:Braibant, M.; Lefevre, P.; de Wit, L.; Peirs, P.; Ooms, J.; Huygen, K.; Andersen, A.B.; Gene 176, 171-176, 1996
A:Title: A Mycobacterium tuberculosis gene cluster encoding proteins of a phosphate trans
A:Reference number: JCS100; MUID:97075926; PMID:8918249
A:Accession: JCS103
A:Molecule type: DNA
A:Residues: 1-374 <BRA>
A:Cross-references: GB:M30046; NID:G149987; PIDN:AAA25374.1; PID:G149988
A:Note: neither the complete nucleic acid sequence nor the complete translation are show
A:Note: the authors translated the initiation codon GTG for residue 1 as Met
R:Andersen, A.B.; Hansen, E.B.
Infect. Immun. 57, 2481-2489, 1989
A:Title: Structure and mapping of antigenic domains of protein antigen b, a 38,000-molecu
A:Reference number: A42930; MUID:9307568; PMID:2545626
A:Accession: A42930
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-374 <AND1>
A:Cross-references: GB:M30046; NID:G149987; PIDN:AAA25374.1; PID:G149988
R:Chang, Z.; Choudhary, A.; Lathigra, R.; Quijcho, F.A.
J. Biol. Chem. 269, 1956-1958, 1994
A:Title: The immunodominant 38-kDa lipoprotein antigen of Mycobacterium tuberculosis is
A:Reference number: A49721; MUID:94124544; PMID:8294447
A:Accession: A49721
A:Molecule type: protein
A:Residues: 25-34 <CHA>
R:Andersen, A.B.; Ljungqvist, L.; Olsen, M.
J. Gen. Microbiol. 136, 477-480, 1990
A:Title: Evidence that protein antigen b of Mycobacterium tuberculosis is involved in pho
A:Reference number: A45820; MUID:90362031; PMID:2118164
A:Accession: A45820
A:Molecule type: protein

A:Residues: 'X',27-28,'X',30-34,'XX',37 <AND2>
 A>Note: confirmed presence of normal signal and absence of lipoprotein attachment
 C:Genetics:

A:Gene: PhoS1; pbs

A:Start codon: GTG

C:Superfamily: phosphate-repressible phosphate-binding protein

C:Keywords: surface antigen

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-374/Product: phosphate specific transporter S #status experimental <MAT>

Query Match 100.0%; Score 96; DB 1; Length 374;

Best Local Similarity 100.0%; Pred. No. 3,3e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NFLPDAQSIQAAAAGPASK 20

Db 272 NFLPDAQSIQAAAAGPASK 291

RESULT 2

JN0013

C:Species: Torpedo californica (Pacific electric ray)

C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 31-Mar-2000

C:Accession: JN0013

R:Linial, M.; Miller, K.; Scheller, R.H.

Neuron 2, 1265-1273, 1989

A:Title: VAR-1: an abundant membrane protein from torpedo cholinergic synaptic vesicles.

A:Reference number: JN0013; MUID:90166593; PMID:2483112

A:Accession: JN0013

A:Molecule type: mRNA

A:Residues: 1-379 <LIN>

C:Comment: Synaptic vesicles are responsible for regulating the storage and release of H

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

C:Keywords: Glycoprotein; membrane protein; phosphoprotein

F:52-350/Domain: long-chain alcohol dehydrogenase homology <LADH>

F:67,127,147/Binding site: carboxylate (Asn) (covalent) #status predicted

F:273/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match

Best Local Similarity 46.98; Score 45; DB 2; Length 379;

Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 FLPLPDAQSIQAAA 15

Db 130 FLPLPDAQSIQAAA 143

RESULT 3

B83370

C:Species: Pseudomonas aeruginosa (strain PA01)

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: B83370

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.D.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: B83370

A:Molecule type: DNA

A:Residues: 1-319 <STO>

A:Cross-references: GB:AE004647; GB:AE004091; NID:99948226; PIDN:AAG05599.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA2211

Query Match

Best Local Similarity 45.88; Score 44; DB 2; Length 319;

Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 5 PDAQSIQAAAAGFA 18

Db 105 PDAEVLAAQAGFA 118

RESULT 4

Tl3154

C:Species: Drosophila melanogaster

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: Tl3154

R:Stankunas, K.; Berger, J.; Ruse, C.; Sinclair, D.A.; Randazzo, P.; Brock, H.W.

Development 125, 4055-4066, 1998

A:Title: The enhancer of polycomb gene of Drosophila encodes a chromatin protein conserv

A:Reference number: Z17611; MUID:98407961; PMID:9735366

A:Accession: Tl3154

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2023 <STA>

A:Cross-references: EMBL:AF079764; NID:93757889; PID:93757890; PIDN:AAC64271.1

A:Experimental source: imaginal disc

C:Genetics:

A:Gene: E(Pc)

A:Cross-references: FlyBase:FBgn0000581

A:Map position: 2

Query Match 45.88; Score 44; DB 2; Length 2023;

Best Local Similarity 56.28; Pred. No. 97;

Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 NFLPDAQSIQAAAAG 16

Db 1777 NFTAASELQAAAAG 1792

RESULT 5

AD1016

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A>Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AD1016

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AD1016

A:Molecule type: DNA

A:Residues: 1-327 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD09229.1; PID:gl6505233; GSPDB:GN00176

C:Genetics:

A:Gene: SIV4441

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

C:Keywords: oxidoreductase

Query Match

Best Local Similarity 44.88; Score 43; DB 2; Length 327;

Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 ILPDAQSIQAAAAGF 17

Db 108 ILPDAISFEQAAASF 122

RESULT 6

AG2710

C:Species: Agrobacterium tumefaciens (strain C58,

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C;Accession: AG2710
 R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A;Reference number: AB2577; MUID:21608550; PMID:11743193
 A;Accession: AG2710
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-191 <KUR>
 A;Cross-references: GB:AE008688; PIDN:AAL42101.1; PID:gl7739483; GSPDB:GN00186
 A;Experimental source: strain C58 (Dupont)
 C;Genetics:
 A;Gene: csnI
 A;Map position: circular chromosome

Query Match 43.8%; Score 42; DB 2; Length 191;
 Best Local Similarity 53.3%; Pred. No. 18;
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 NFFLLPDAQSIQAAAA 15
 ||||| :||| :
 Db 28 NFFLLPQKALRANAA 42

RESULT 7
 H97492
 cultivar specific nodulation protein csnI (AF185636) [imported] - Agrobacterium tumefaci
 C;Species: Agrobacterium tumefaciens
 C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
 C;Accession: H97492
 R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A;Reference number: A97359; MUID:21608551; PMID:11743194
 A;Accession: H97492
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-191 <KUR>
 A;Cross-references: GB:AE007869; PIDN:AAK86897.1; PID:g15156121; GSPDB:GN00169
 C;Genetics:
 A;Gene: AGR C 2015
 A;Map position: circular chromosome

Query Match 43.8%; Score 42; DB 2; Length 191;
 Best Local Similarity 53.3%; Pred. No. 18;
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 NFFLLPDAQSIQAAAA 15
 ||||| :||| :
 Db 28 NFFLLPQKALRANAA 42

RESULT 8
 H84216
 hypothetical protein vng0587h [imported] - Halobacterium sp. NRC-1
 C;Species: Halobacterium sp. NRC-1
 C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C;Accession: H84216
 R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
 ; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
 A;Title: Genome sequence of Halobacterium species NRC-1
 A;Reference number: A84160; MUID:20504483; PMID:11016950
 A;Accession: H84216
 A;Status: preliminary
 A;Molecule type: DNA

A;Residues: 1-303 <STO>
 A;Cross-references: GB:AE004437; MID:g10580182; PIDN:AAG19100.1; GSPDB:GN00138
 C;Genetics:
 A;Gene: VNG0587H

Query Match 43.8%; Score 42; DB 2; Length 303;
 Best Local Similarity 50.0%; Pred. No. 29;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 LPDQAQSIQAAAAAGFAS 19
 ||||| :||| :
 Db 158 LPDPQEIPEASAGFGT 173

RESULT 9
 AI0529
 probable membrane protein yael [imported] - Salmonella enterica subsp. enterica serovar 7
 C;Species: Salmonella enterica subsp. enterica serovar Typhi
 A;Note: this species has also been called Salmonella typhi
 C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C;Accession: AI0529
 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
 A;Reference number: AB0502; MUID:21534947; PMID:11677608
 A;Accession: AI0529
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-450 <PAR>
 A;Cross-references: GB:AL513382; PIDN:CAD06681.1; PID:g16501504; GSPDB:GN00176
 C;Genetics:
 A;Gene: yael
 C;Superfamily: Escherichia coli probable zinc proteinase yael

Query Match 43.8%; Score 42; DB 2; Length 450;
 Best Local Similarity 50.0%; Pred. No. 44;
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 LLPDQAQSIQAAAAAGFA 18
 ||||| :||| :
 Db 287 LTPDTKSVNGKAEFGA 302

RESULT 10
 A47022
 chitinase (EC 3.2.1.14) I - Rhizopus oligosporus
 C;Species: Rhizopus oligosporus
 C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 29-Oct-1999
 C;Accession: A47022; S27418
 R;Yanai, K.; Takaya, N.; Kojima, N.; Horiuchi, H.; Ohta, A.; Takagi, M.
 J. Bacteriol. 174, 7398-7406, 1992
 A;Title: Purification of two chitinases from Rhizopus oligosporus and isolation and sequ
 A;Reference number: A47022; MUID:93054356; PMID:1429462
 A;Accession: A47022
 A;Status: preliminary
 A;Molecule type: DNA; protein
 A;Residues: 1-540 <YAN>
 A;Cross-references: EMBL:D10157; MID:g218026; PIDN:BAA01021.1; PID:d1001490; PID:g218027
 A;Note: sequence extracted from NCBI backbone (NCBIP:118192)
 C;Genetics:
 A;Gene: chII
 C;Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 43.8%; Score 42; DB 2; Length 540;
 Best Local Similarity 50.0%; Pred. No. 53;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 PDAQSIQAAAAAGFASK 20
 ||||| :||| :
 Db 415 PSASALEIAAASFQSK 430

RESULT 11
 576211
 hypothetical protein slr0442 - *Synechocystis* sp. (strain PCC 6803)
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C:Accession: S76211
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S76211
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-611 <KAN>
 A:Cross-references: EMBL:D90914; GB:AB001339; NID:G1653477; PIDN:BAAL8470.1; PID:dl010200
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 Query Match 43.8%; Score 42; DB 2; Length 611;
 Best Local Similarity 56.2%; Pred. No. 60;
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 4 LPLDPAQSIQAAAAGFAS 19
 DB 282 IDLLNSIQNAAGGFPA 297

RESULT 12
 A39596
 progesterone receptor B form - mouse
 C:Species: *Mus musculus* (house mouse)
 C:Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 20-Aug-1999
 C:Accession: A39596; I49111
 R:Schott, D.R.; Shiyanala, G.; Schneider, W.; Parry, G.
 Biochemistry 30, 7014-7020, 1991
 A:Title: Molecular cloning, sequence analyses, and expression of complementary DNA encoding progesterone receptor B form
 A:Reference number: A39596; MUID:9129759; PMID:2069958
 A:Accession: A39596
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-923 <SCH>
 A:Cross-references: GB:M68915; GB:J05333; NID:G200471; PIDN:AAA39971.1; PID:G200472
 R:Hagihara, K.; Wu-Peng, X.S.; Funabashi, T.; Kato, J.; Pfaff, D.W.
 Biochem. Biophys. Res. Commun. 205, 1093-1101, 1994
 A:Title: Nucleic acid sequence and DNase hypersensitive sites of the 5' region of the mouse progesterone receptor B form
 A:Reference number: I49111; MUID:95100931; PMID:7802637
 A:Accession: I49111
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-9 <HAG>
 A:Cross-references: EMBL:U12644; NID:9639916; PIDN:AAA66067.1; PID:G639917
 C:Superfamily: progesterone receptor; erba transforming protein homology
 C:Keywords: DNA binding; nucleus; steroid hormone receptor; transcription regulation; zinc finger
 F:555-819/Domain: erba transforming protein homology <ERBA>
 F:557-577/Region: zinc finger
 F:593-617/Region: zinc finger
 Query Match 43.8%; Score 42; DB 2; Length 923;
 Best Local Similarity 42.1%; Pred. No. 93;
 Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
 QY 1 NPLPDAQSIQAAAAGFAS 19
 DB 533 NYLRPDSASQSPQVGFD 551

RESULT 13
 153280
 progesterone receptor B form - rat
 C:Species: *Rattus norvegicus* (Norway rat)
 C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 20-Aug-1999
 C:Accession: I53280; B23733; A49574; A23733
 R:Park-Sarge, O.K.; Mayo, K.E.
 Endocrinology 134, 709-718, 1994
 A:Title: Regulation of the progesterone receptor gene by gonadotropins and cyclic adenosine 3',5'-phosphate
 A:Reference number: I53280; MUID:94130817; PMID:8299566
 A:Accession: I53280
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-923 <PAR>
 A:Cross-references: GB:L16922; NID:G463282; PIDN:AAA19516.1; PID:G463283
 R:Park, O.K.; Mayo, K.E.
 Mol. Endocrinol. 5, 967-978, 1991
 A:Title: Transient expression of progesterone receptor messenger RNA in ovarian granulosa cells
 A:Reference number: A23733; MUID:92049379; PMID:1840636
 A:Accession: B23733
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 555-589, W, 591-624, 742-910 <PAR>
 A:Cross-references: GB:S64044
 R:Kraus, W.L.; Montano, M.M.; Katzenellenbogen, B.S.
 Mol. Endocrinol. 7, 1603-1616, 1993
 A:Title: Cloning of the rat progesterone receptor gene 5'-region and identification of two alternative promoters
 A:Reference number: A49574; MUID:94195318; PMID:8145766
 A:Accession: A49574
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-16 <KRA>
 A:Cross-references: GB:S69361; NID:G546178
 C:Superfamily: progesterone receptor; erba transforming protein homology
 C:Keywords: DNA binding; nucleus; steroid hormone receptor; zinc finger
 F:555-619/Domain: erba transforming protein homology <ERBA>
 F:557-577/Region: zinc finger
 F:593-617/Region: zinc finger
 Query Match 43.8%; Score 42; DB 2; Length 923;
 Best Local Similarity 42.1%; Pred. No. 93;
 Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
 QY 1 NPLPDAQSIQAAAAGFAS 19
 DB 533 NYLRPDSASQSPQVGFD 551

RESULT 14
 F83333
 hypothetical protein PA2504 [imported] - *Pseudomonas aeruginosa* (strain PA01)
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: F83333
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Brj
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: F83333
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-205 <STO>
 A:Cross-references: GB:AE004677; GB:AE004091; NID:99948548; PIDN:AA05892.1; GSPDB:GN0011
 A:Experimental source: strain PA01
 A:Genetics:
 A:Gene: PA2504
 Query Match 42.7%; Score 41; DB 2; Length 205;
 Best Local Similarity 47.4%; Pred. No. 29;
 Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
 QY 2 FLLPDAQSIQAAAAGFASK 20
 DB 20 FYKPDATIDRVAGESVK 38

RESULT 15

H97124
 hypothetical protein CAC1822 [imported] - Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: H97124
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: H97124
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-281 <KUR>
 A:Cross-references: GB:AF001437; PIDN:AAK79787.1; PID:g15024797; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC1822

Query Match 42.7%; Score 41; DB 2; Length 281;
 Best Local Similarity 53.3%; Pred.No. 40;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 PDAQSIQAAAAGFAS 19
 |||::|||:
 Db 137 PDVQSLKAAENLAN 151

Search completed: March 10, 2004, 12:12:25
 Job time : 11.7692 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:50:51 ; Search time 47.5641 Seconds
(without alignments)
118.807 Million cell updates/sec

Title: US-10-044-703-47

Perfect score: 96

Sequence: 1 NPLLPDAQSIQAAAAGFASK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96	100.0	20	4	AAE12262 Mycobacte
2	96	100.0	351	6	ABU63550 Mature My
3	96	100.0	373	6	ABU63549 Mycobacte
4	96	100.0	374	2	AAR30090 38 kDa li
5	96	100.0	374	2	AAW64363 Mycobacte
6	96	100.0	374	2	AAW64364 Mycobacte
7	96	100.0	374	2	AAW81730 M. tuberc
8	96	100.0	374	2	AAW81731 M. tuberc
9	96	100.0	374	2	AAW81731 M. tuberc
10	96	100.0	374	2	AAW81731 M. tuberc
11	96	100.0	374	2	AAW81731 M. tuberc
12	96	100.0	374	2	AAW81731 M. tuberc
13	96	100.0	374	2	AAW81731 M. tuberc
14	96	100.0	374	2	AAW81731 M. tuberc
15	96	100.0	374	2	AAW81731 M. tuberc
16	96	100.0	374	2	AAW81731 M. tuberc
17	96	100.0	374	2	AAW81731 M. tuberc
18	96	100.0	374	2	AAW81731 M. tuberc
19	96	100.0	374	2	AAW81731 M. tuberc
20	96	100.0	374	2	AAW81731 M. tuberc
21	96	100.0	374	2	AAW81731 M. tuberc
22	96	100.0	374	2	AAW81731 M. tuberc
23	96	100.0	374	2	AAW81731 M. tuberc
24	96	100.0	374	2	AAW81731 M. tuberc
25	96	100.0	374	2	AAW81731 M. tuberc

26	96	100.0	802	2	AAV39176 M. tuberc
27	96	100.0	802	2	AAV39224 M. tuberc
28	96	100.0	802	5	AAU74592 Antigene
29	96	100.0	983	4	AAU01901 M. tuberc
30	90	93.8	23	4	AAE12273 Mycobacte
31	55	57.3	12	4	AAE12257 Mycobacte
32	44	45.8	9	2	AAE12257 Mycobacte
33	44	45.8	2023	4	AAE12257 Mycobacte
34	44	45.8	7257	3	AAE12257 Mycobacte
35	43	44.8	106	3	AAE12257 Mycobacte
36	43	44.8	220	4	AAE12257 Mycobacte
37	43	44.8	235	6	AAE12257 Mycobacte
38	43	44.8	280	4	AAE12257 Mycobacte
39	43	44.8	280	4	AAE12257 Mycobacte
40	42.5	44.3	128	4	AAE12257 Mycobacte
41	42	43.8	161	7	AAE12257 Mycobacte
42	42	43.8	302	4	AAE12257 Mycobacte
43	42	43.8	388	5	AAE12257 Mycobacte
44	42	43.8	422	2	AAE12257 Mycobacte
45	42	43.8	540	2	AAE12257 Mycobacte

ALIGNMENTS

RESULT 1

AAE12262
ID AAE12262 standard; peptide; 20 AA.

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

XX AC AAE12262;

Query Match 100.0%; Score 96; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.2e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFLLPDAQSIQAAAGFASK 20
DB 1 NFLLPDAQSIQAAAGFASK 20

RESULT 2

ABU63550
ID ABU63550 standard; protein; 351 AA.

XX AC ABU63550;

XX DT 25-SEP-2003 (first entry)

XX DE Mature Mycobacterium tuberculosis 38kDa protein.

XX KW Antimicrobial; cytostatic; type 1 response; 38kDa protein;
XX KW T-helper-1 T-cell response; interleukin-12; IL-12;
XX KW cell-mediated immunity; pathogen infection; vaccine;
XX KW interferon-gamma sensitive tumour.

XX OS Mycobacterium tuberculosis.

PH Key Location/Qualifiers
FT Modified-site 1

FT /label= OTHER
FT /note= "OTHER= N-acyl diglyceride cysteine"

XX PN US6517839-B1.

XX PD 11-FEB-2003.

XX PF 17-JUL-1998; 98US-00118426.
XX PR 18-JUL-1997; 97US-0052970P.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Modlin RL, Libraty DH;

XX DR WPI; 2003-553609/52.

XX CC Inducing interleukin-12 and type 1/T-helper-1 T-cell response to
PT stimulate cell-mediated immunity for preventing and treating pathogen
PT infection involves administering lipopeptide having N-terminal ester- or
PT amide-linked fatty acyl group.

XX PS Disclosure; Col 25-26; 27pp; English.

XX CC The invention describes a method of inducing type 1/T-helper-1 T-cell
CC response or interleukin-12 (IL-12) in a cell involves administration of a
CC lipopeptide (I) having an N-terminal ester- or amide-linked fatty acyl
CC group. The methods are used to stimulate cell-mediated immunity for
CC prevention and treatment of pathogen infections (e.g. Mycobacterium
CC tuberculosis) and for treatment of a interferon-gamma sensitive tumour.
CC Also useful in the development of new vaccine and therapeutic strategies.
CC This is the amino acid sequence of mature Mycobacterium tuberculosis
CC 38kDa protein from which lipopeptides can be isolated

XX SQ Sequence 351 AA;

Query Match 100.0%; Score 96; DB 6; Length 351;
Best Local Similarity 100.0%; Pred. No. 3.5e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFLLPDAQSIQAAAGFASK 20
DB 249 NFLLPDAQSIQAAAGFASK 268

RESULT 3

ABU63549

ID ABU63549 standard; protein; 373 AA.

XX AC ABU63549;

XX DT 25-SEP-2003 (first entry)

XX DE Mycobacterium tuberculosis 38kDa protein.

XX KW Antimicrobial; cytostatic; type 1 response; 38kDa protein;
XX KW T-helper-1 T-cell response; interleukin-12; IL-12;
XX KW cell-mediated immunity; pathogen infection; vaccine;
XX KW interferon-gamma sensitive tumour.

XX OS Mycobacterium tuberculosis.

PH Key Location/Qualifiers
FT Modified-site 1

FT /label= OTHER
FT /note= "OTHER= N-acyl diglyceride cysteine"

XX PN US6517839-B1.

XX PD 11-FEB-2003.

XX PF 17-JUL-1998; 98US-00118426.
XX PR 18-JUL-1997; 97US-0052970P.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Modlin RL, Libraty DH;

XX DR WPI; 2003-553609/52.

XX CC Inducing interleukin-12 and type 1/T-helper-1 T-cell response to
PT stimulate cell-mediated immunity for preventing and treating pathogen
PT infection involves administering lipopeptide having N-terminal ester- or
PT amide-linked fatty acyl group.

XX PS Disclosure; Col 23-24; 27pp; English.

XX CC The invention describes a method of inducing type 1/T-helper-1 T-cell
CC response or interleukin-12 (IL-12) in a cell involves administration of a
CC lipopeptide (I) having an N-terminal ester- or amide-linked fatty acyl
CC group. The methods are used to stimulate cell-mediated immunity for
CC prevention and treatment of pathogen infections (e.g. Mycobacterium
CC tuberculosis) and for treatment of a interferon-gamma sensitive tumour.
CC Also useful in the development of new vaccine and therapeutic strategies.
CC This is the amino acid sequence of Mycobacterium tuberculosis 38kDa
CC protein from which lipopeptides can be isolated

XX SQ Sequence 373 AA;

Query Match 100.0%; Score 96; DB 6; Length 373;
Best Local Similarity 100.0%; Pred. No. 3.8e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFLLPDAQSIQAAAGFASK 20
DB 271 NFLLPDAQSIQAAAGFASK 290

RESULT 4

AAR30090
ID AAR30090 standard; protein; 374 AA.

XX AC AAR30090;

XX DT 25-MAR-2003 (revised)

XX DT 05-MAY-1993 (first entry)

XX DE 38 kDa lipoprotein antigen.

D5 272 NFLLPDAQSIQAAAAAGFASK 291

RESULT 6

AAW64364
ID AAW64364 standard; protein; 374 AA.

AC AAW64364;
XX

DT 09-NOV-1998 (first entry)
XX

DE Mycobacterium 39 kDa antigen.
XX

KW Tuberculosis; infection; diagnosis; 38 kDa antigen.
XX

OS Mycobacterium tuberculosis.
XX

PN WO9816645-A2.
XX

PD 23-APR-1998.
XX

PF 07-OCT-1997; 97WO-US018214.
XX

PR 11-OCT-1996; 96US-00729622.
XX

PR 13-MAR-1997; 97US-00818111.
XX

PA (CORI-) CORIXA CORP.
XX

PI Read SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX

PI Vedvick TS, Twardzik DR, Lodes MJ;
XX

DR WPI; 1998-251292/22.
XX

DR N-PSDB; AAW44414.
XX

PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
develop products for the detection of M. tuberculosis infection and
diagnosis of tuberculosis.
XX

PS Disclosure; Page 168-169; 250pp; English.
XX

CC This polypeptide comprises Mycobacterium tuberculosis 38 kDa antigen. A
TbRa3-38 kD-Tb38-1 fusion protein (see AAW64364) has been constructed
that can be used in the serodiagnosis of tuberculosis. The invention
relates to compositions and methods for diagnosing tuberculosis. It
provides polypeptides (see AAW64291-W64379) comprising an antigenic
portion of a soluble M. tuberculosis antigen, an immunogenic portion of
an M. tuberculosis antigen, or a fusion protein, as well as DNA sequences
encoding such polypeptides, recombinant expression vectors and
transformed or transfected host cells. Also claimed are methods and
diagnostic kits for detecting M. tuberculosis infection in a patient
XX

SQ Sequence 374 AA;

Query Match 100.0%; Score 96; DB 2; Length 374;

Best Local Similarity 100.0%; Pred. No. 3.8e-08; Mismatches 0; Gaps 0;

Matches 20; Conservative 0; Indels 0; Gaps 0;

QY 1 NFLLPDAQSIQAAAAAGFASK 20

DB 272 NFLLPDAQSIQAAAAAGFASK 291

RESULT 7

AAW81730

ID AAW81730 standard; protein; 374 AA.

AC AAW81730;
XX

DT 27-JAN-1999 (first entry)
XX

DE M. tuberculosis fusion protein TbRa3/38kD/Tb38-1.
XX

KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
XX

KW vaccine; pharmaceutical; infection; diagnosis.

XX Synthetic.

OS Mycobacterium tuberculosis.

PN WO9816646-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US018293.

XX 11-OCT-1996; 96US-00730510.

XX 13-MAR-1997; 97US-00818112.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

XX Vedvick TS, Twardzik DR, Lodes MJ;

XX WPI; 1998-261042/23.

XX N-PSDB; AAW64522.

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
develop products for the detection of M. tuberculosis infection and for
diagnosis, treatment and prevention of tuberculosis.

PS Claim 37; Page 156-157; 230pp; English.

CC This sequence represents a fusion protein containing the immunogenic
polypeptides TbRa3, 38kD and Tb38-1 from Mycobacterium tuberculosis (MT).
This fusion protein can be used in a method for inducing protective
immunity against tuberculosis (TB). This sequence can be formulated into
vaccines and/or pharmaceutical compositions for immunising against M.
tuberculosis infection or may be used for the diagnosis of TB

SQ Sequence 374 AA;

Query Match 100.0%; Score 96; DB 2; Length 374;

Best Local Similarity 100.0%; Pred. No. 3.8e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFLLPDAQSIQAAAAAGFASK 20

DB 272 NFLLPDAQSIQAAAAAGFASK 291

RESULT 8

AAW81731

ID AAW81731 standard; protein; 374 AA.

XX AAW81731;

XX 27-JAN-1999 (first entry)

XX M. tuberculosis 38kD antigen protein.

XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;

KW vaccine; pharmaceutical; infection; diagnosis.

XX Mycobacterium tuberculosis.

XX WO9816646-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US018293.

XX 11-OCT-1996; 96US-00730510.

XX 13-MAR-1997; 97US-00818112.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

XX Vedvick TS, Twardzik DR, Lodes MJ;

PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX WPI; 1998-261042/23.
 DR N-PSDB; AAV64523.
 XX
 PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and for
 PT diagnosis, treatment and prevention of tuberculosis.
 XX
 PS Claim 23; Page 159-160; 230pp; English.
 XX
 CC This sequence represents a 38kD antigen from Mycobacterium tuberculosis
 CC (MT). This antigen can be used in a method for inducing protective
 CC immunity against tuberculosis (TB). This sequence can be formulated into
 CC vaccines and/or pharmaceutical compositions for immunising against M.
 CC tuberculosis infection or may be used for the diagnosis of TB
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 96; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3.8e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NFLLPDAQSIQAAAAGFASK 20
 |||||
 Db 272 NFLLPDAQSIQAAAAGFASK 291
 |||||
 RESULT 9
 AAY32061
 ID AAY32061 standard; protein; 374 AA.
 AC AAY32061;
 XX
 DT 17-JAN-2000 (first entry)
 DE Mycobacterium tuberculosis antigen Tbra3-38kD-Tb38-1 fusion.
 KW Tuberculosis; antigen; fusion protein; Tbra3; 38kD; Tb38-1; diagnosis;
 KW therapy; vaccine; immunogen.
 XX Mycobacterium tuberculosis.
 OS WO9951748-A2.
 PN 14-OCT-1999.
 PD 07-APR-1999; 99WO-US007717.
 PF 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.
 XX (CORI-) CORIXA CORP.
 PA Skeiky YAW, Alderson M, Campos-Neto A;
 PI WPI; 1999-601610/51.
 DR N-PSDB; AAZ20196.
 XX
 PT New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis.
 XX
 PS Claim 1; Fig 3A-D; 83pp; English.
 XX
 CC This sequence represents a recombinant Mycobacterium tuberculosis tri-
 CC antigen fusion protein composed of the antigens Tbra3, 39kD and Tb38-1.
 CC The fusion protein is expressed in host cells using a vector carrying a
 CC polynucleotide (see AAZ20196) comprising the 3 coding sequences for the
 CC antigens. The invention provides fusion proteins (see AAY32059-71)
 CC containing at least 2 M. tuberculosis antigens. The new fusion proteins
 CC and polynucleotides encoding them are useful as vaccines for preventing
 CC tuberculosis (claimed), for diagnosis (via in vitro assays or intradermal
 CC skin tests for detection of anti-M. tuberculosis antibodies), monitoring

CC of disease progression, and treatment of tuberculosis. They are more
 CC effective immunogens than mixtures of the individual protein components
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 96; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3.8e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NFLLPDAQSIQAAAAGFASK 20
 |||||
 Db 272 NFLLPDAQSIQAAAAGFASK 291
 |||||
 RESULT 10
 AAY39018
 ID AAY39018 standard; protein; 374 AA.
 XX
 AC AAY39018;
 XX
 DT 05-NOV-1999 (first entry)
 DE M. tuberculosis 38 kDa antigen protein.
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 XX Mycobacterium tuberculosis.
 OS WO9942118-A2.
 PN 26-AUG-1999.
 PD 17-FEB-1999; 99WO-US003265.
 PF 18-FEB-1998; 98US-00024753.
 PR 05-MAY-1998; 98US-00072596.
 XX (CORI-) CORIXA CORP.
 PA Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX WPI; 1999-527416/44.
 DR N-PSDB; AAZ19112.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis.
 PS Claim 53; Page 199-200; 323pp; English.
 XX
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against M.
 CC tuberculosis infection. The new detection methods are needed as current
 CC vaccination strategies do not provide 100% immunity
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 96; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3.8e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NFLLPDAQSIQAAAAGFASK 20
 |||||
 Db 272 NFLLPDAQSIQAAAAGFASK 291
 |||||
 RESULT 11
 AAY39017
 ID AAY39017 standard; protein; 374 AA.
 XX

```

AC AAY39017;
AD 05-NOV-1999 (first entry)
DE M. tuberculosis fusion protein TbrA3/38kDa/Tb38-1.
XX Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity.
XX Mycobacterium tuberculosis.
OS WO9942118-A2.
PN 26-AUG-1999.
PD 17-FEB-1999; 99WO-US003265.
PF 18-FEB-1998; 98US-00024753.
PR 05-MAY-1998; 98US-00072596.
XX (CORI-) CORIXA CORP.
PA Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX WPI; 1999-527416/44.
DR N-PSDB; AAZ19111.
XX New polypeptide comprising antigenic portions of M. tuberculosis.
PT Example 10; Page 196-198; 323pp; English.
PS This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against M.
CC tuberculosis infection. The new detection methods are needed as current
CC vaccination strategies do not provide 100% immunity
XX
XX Sequence 374 AA;
XX
Query Match 100.0%; Score 96; DB 2; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.8e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NFLLPDAQSIQAAAAGFASK 20
DB 272 NFLLPDAQSIQAAAAGFASK 291
RESULT 12
AAY39161
ID AAY39161 standard; protein; 374 AA.
XX
AC AAY39161;
XX
XX 05-NOV-1999 (first entry)
XX
XX M. tuberculosis antigen 38 kD amino acid sequence.
DE
XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
XX Mycobacterium tuberculosis.
OS
XX WO9942076-A2.
PN
XX 26-AUG-1999.
PD
XX 17-FEB-1999; 99WO-US003268.
PF
XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
XX Mycobacterium tuberculosis.
OS
XX WO9942076-A2.
PN
XX 26-AUG-1999.
PD
XX 17-FEB-1999; 99WO-US003268.
PF
XX
XX

```

```

PR 18-FEB-1998; 98US-00025197.
PR 05-MAY-1998; 98US-00072967.
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX WPI; 1999-527409/44.
DR N-PSDB; AAZ19324.
XX
XX New antigens from Mycobacterium tuberculosis useful in diagnostic skin
PT tests and protective or therapeutic vaccines or compositions.
XX Claim 23; Page 154-155; 299pp; English.
XX
XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
CC polypeptides fragments, can be used in pharmaceutical compositions or
CC vaccines to generate a protective or therapeutic immune response to M.
CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
CC killer cells and/or macrophages in tuberculosis-immune subjects. AAZ19249
CC to AAZ19460 and AAY39083 to AAY39225 are used in the exemplification of
CC the present invention
XX
XX Sequence 374 AA;
XX
Query Match 100.0%; Score 96; DB 2; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.8e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NFLLPDAQSIQAAAAGFASK 20
DB 272 NFLLPDAQSIQAAAAGFASK 291
RESULT 13
AAY39160
ID AAY39160 standard; protein; 374 AA.
XX
AC AAY39160;
XX
XX 05-NOV-1999 (first entry)
XX
XX M. tuberculosis TbrA3, 38 kD and Tb38-1 fusion protein.
DE
XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
XX Synthetic.
OS
XX Mycobacterium tuberculosis.
OS
XX WO9942076-A2.
PN
XX 26-AUG-1999.
PD
XX 17-FEB-1999; 99WO-US003268.
PF
XX 18-FEB-1998; 98US-00025197.
PR 05-MAY-1998; 98US-00072967.
XX
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX WPI; 1999-527409/44.
DR N-PSDB; AAZ19323.
XX
XX

```

PT New antigens from Mycobacterium tuberculosis useful in diagnostic skin
 PT tests and protective or therapeutic vaccines or compositions.
 XX
 PS Claim 37; Page 151-152; 299pp; English.
 XX
 CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
 CC vaccines and fusion protein containing M. tuberculosis Ag's. M. M.
 CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
 CC polypeptides fragments, can be used in pharmaceutical compositions or
 CC vaccines to generate a protective or therapeutic immune response to M.
 CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
 CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
 CC killer cells and/or macrophages in tuberculosis-immune subjects. AA219249
 CC to AA219460 and AA239083 to AA239225 are used in the exemplification of
 CC the present invention
 XX
 XX Sequence 374 AA;
 Query Match 100.0%; Score 96; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3.8e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NFLLPDAQSIQAAAAGFASK 20
 DB 272 NFLLPDAQSIQAAAAGFASK 291
 RESULT 14
 AA211840
 ID AA211840 standard; protein; 374 AA.
 AC AA211840;
 DT 18-DEC-2001 (first entry)
 DE Mycobacterium tuberculosis 38kDa-LP protein.
 XX 38kDa-LP protein; antibiotic; vaccine; CD8 T-cell; epitope; medicament;
 XX prophylactic; therapeutic; mycobacterial infection.
 XX Mycobacterium tuberculosis.
 XX WO200170764-A2.
 XX 27-SEP-2001.
 XX 20-MAR-2001; 2001WO-GB001205.
 XX 20-MAR-2000; 2000GB-00006692.
 XX (GLAX) GLAXO GROUP LTD.
 XX Klein MR;
 XX WPI; 2001-607517/69.
 XX N-PSDB; AAD18885.
 XX Use of polypeptide comprising CD8 T-cell epitopes derived from
 PT Mycobacterium tuberculosis proteins in the manufacture of a medicament
 PT for vaccinating prophylactically or therapeutically against mycobacterial
 PT infection.
 XX Claim 3; Page 50-52; 54pp; English.
 XX
 CC The invention relates to use of polypeptides comprising CD8 T-cell
 CC epitopes derived from Mycobacterium tuberculosis proteins. The invention
 CC is useful in the manufacture of a medicament for prophylactic or
 CC therapeutic vaccination against mycobacterial infection. The polypeptide
 CC and the expression vector are capable of stimulating CD8 T-cell response.
 CC The invention also provides a vaccine composition which comprises
 CC polypeptide or expression vector useful for vaccinating a pre-selected
 CC host against a mycobacterial infection. The invention further relates to

CC a method for determining the presence or absence of CD8 T-cell response
 CC to epitope sequence by identifying the presence or absence of T-cell that
 CC recognise the epitope sequence in a sample from the host. The present
 CC sequence is Mycobacterium tuberculosis 38kDa-LP protein related to the
 CC invention
 XX
 XX Sequence 374 AA;
 Query Match 100.0%; Score 96; DB 4; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3.8e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NFLLPDAQSIQAAAAGFASK 20
 DB 272 NFLLPDAQSIQAAAAGFASK 291
 RESULT 15
 AA229719
 ID AA229719 standard; protein; 374 AA.
 AC AA229719;
 DT 27-JAN-2003 (first entry)
 DE Mycobacterium sp. 38kd antigenic protein.
 XX Vaccine; immunity; diagnostic agent; gene therapy; 38kd antigen.
 XX Mycobacterium sp.
 XX Key Location/Qualifiers
 FT Active-site 1..2
 FT Binding-site 1..2
 FT Modified-site 1..2
 FT Cleavage-site 1..2
 FT Cross-links 1..2
 FT Disulfide-bond 1..2
 FT Domain 1..2
 FT Misc-difference 1..2
 FT /note= "Encoded by "
 XX WO200272792-A2.
 XX 19-SEP-2002.
 XX 13-MAR-2002; 2002WO-US008223.
 XX 13-MAR-2001; 2001US-0275837P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Brannon M, Guderian J;
 XX WPI; 2002-759844/82.
 XX N-PSDB; AAD47096.
 XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.
 XX Disclosure; Page 109-110; 155pp; English.
 XX
 CC The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic

CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides, as
CC in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is
CC Mycobacterium sp. 38kD antigenic protein

XX

SQ Sequence 374 AA;

Query Match 100.0%; Score 96; DB 5; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.8e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 NFLLPDAQSIQAAAAGPASK 20

Db 272 NFLLPDAQSIQAAAAGPASK 291

Search completed: March 10, 2004, 12:05:29
Job time : 49.5641 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 12:11:07 ; Search time 21.359 Seconds
(without alignments)
168.061 Million cell updates/sec

Title: US-10-044-703-49

Perfect score: 81
Sequence: 1 TGSAGAGIAQAAGTVNI 17

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 21113259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81	100.0	17	9	US-09-813-333-49
2	81	100.0	17	13	US-10-044-703-49
3	81	100.0	18	9	US-09-813-333-55
4	81	100.0	18	13	US-10-044-703-55
5	81	100.0	374	9	US-09-287-849-6
6	81	100.0	374	9	US-09-287-849-40
7	81	100.0	374	14	US-10-193-002-148
8	81	100.0	374	14	US-10-193-002-150
9	81	100.0	374	14	US-10-084-843-153
10	81	100.0	374	14	US-10-084-843-155
11	81	100.0	374	14	US-10-359-460-6
12	81	100.0	374	14	US-10-359-460-40
13	81	100.0	374	14	US-10-098-732A-39
14	81	100.0	652	14	US-10-193-002-350
15	81	100.0	652	14	US-10-084-843-355

16	81	100.0	802	9	US-09-287-849-10	Sequence 10, Appl
17	81	100.0	802	14	US-10-193-002-209	Sequence 209, App
18	81	100.0	802	14	US-10-193-002-346	Sequence 346, App
19	81	100.0	802	14	US-10-084-843-214	Sequence 214, App
20	81	100.0	802	14	US-10-084-843-351	Sequence 351, App
21	81	100.0	802	14	US-10-359-460-10	Sequence 10, Appl
22	50	61.7	2159	14	US-10-123-155-87	Sequence 87, Appl
23	50	61.7	2159	14	US-10-146-731-87	Sequence 87, Appl
24	50	61.7	2159	14	US-10-140-472-87	Sequence 87, Appl
25	50	61.7	2159	14	US-10-141-761-87	Sequence 87, Appl
26	50	61.7	2159	14	US-10-142-885-87	Sequence 87, Appl
27	50	61.7	2159	14	US-10-158-790-87	Sequence 87, Appl
28	50	61.7	2159	15	US-10-137-871-87	Sequence 87, Appl
29	50	61.7	2159	15	US-10-140-923-87	Sequence 87, Appl
30	50	61.7	2159	15	US-10-141-756-87	Sequence 87, Appl
31	50	61.7	2159	15	US-10-141-759-87	Sequence 87, Appl
32	50	61.7	2159	15	US-10-140-805-87	Sequence 87, Appl
33	50	61.7	2159	15	US-10-140-864-87	Sequence 87, Appl
34	48	59.3	957	14	US-10-184-644-603	Sequence 603, App
35	48	59.3	957	14	US-10-184-634-603	Sequence 603, App
36	48	59.3	957	14	US-10-123-155-513	Sequence 513, App
37	48	59.3	957	14	US-10-146-731-513	Sequence 513, App
38	48	59.3	957	14	US-10-140-472-513	Sequence 513, App
39	48	59.3	957	14	US-10-141-761-513	Sequence 513, App
40	48	59.3	957	14	US-10-142-885-513	Sequence 513, App
41	48	59.3	957	14	US-10-158-790-513	Sequence 513, App
42	48	59.3	957	15	US-10-137-871-513	Sequence 513, App
43	48	59.3	957	15	US-10-140-923-513	Sequence 513, App
44	48	59.3	957	15	US-10-141-756-513	Sequence 513, App
45	48	59.3	957	15	US-10-141-759-513	Sequence 513, App

ALIGNMENTS

RESULT 1

US-09-813-333-49
; Sequence 49, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/09/813.333
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 49
; TYPE: PRT
; LENGTH: 17
; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-49

Query Match 100.0%; Score 81; DB 9; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.8e-05; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0

QY 1 TGSAGAGIAQAAGTVNI 17
|||||
Db 1 TGSAGAGIAQAAGTVNI 17

RESULT 2

US-10-044-703-49
; Sequence 49, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703

; CURRENT FILING DATE: 2002-05-20
 ; PRIOR APPLICATION NUMBER: 60/190,834
 ; PRIOR FILING DATE: 2000-03-20
 ; NUMBER OF SEQ ID NOS: 81
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 49
 ; LENGTH: 17
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 US-10-044-703-49

Query Match 100.0%; Score 81; DB 13; Length 17;
 Best Local Similarity 100.0%; Pred. No. 2.8e-05;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSGAGIAQAAAGTVNI 17
 |||||
 Db 1 TGSGAGIAQAAAGTVNI 17

RESULT 3
 US-09-813-333-55
 ; Sequence 55, Application US/09813333
 ; Patent No. US20020119160A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DeGroot, Anne S
 ; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
 ; FILE REFERENCE: 17999-004 US
 ; CURRENT APPLICATION NUMBER: US/09/813,333
 ; CURRENT FILING DATE: 2001-03-20
 ; PRIOR APPLICATION NUMBER: 60/190,834
 ; PRIOR FILING DATE: 2000-03-20
 ; NUMBER OF SEQ ID NOS: 81
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 55
 ; LENGTH: 18
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 US-09-813-333-55

Query Match 100.0%; Score 81; DB 9; Length 18;
 Best Local Similarity 100.0%; Pred. No. 3e-05;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSGAGIAQAAAGTVNI 17
 |||||
 Db 1 TGSGAGIAQAAAGTVNI 17

RESULT 4
 US-10-044-703-55
 ; Sequence 55, Application US/10044703
 ; Publication No. US20020192233A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DeGroot, Anne S
 ; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
 ; FILE REFERENCE: 17999-004 US
 ; CURRENT APPLICATION NUMBER: US/10/044,703
 ; CURRENT FILING DATE: 2002-05-20
 ; PRIOR APPLICATION NUMBER: 60/190,834
 ; PRIOR FILING DATE: 2000-03-20
 ; NUMBER OF SEQ ID NOS: 81
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 55
 ; LENGTH: 18
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 US-10-044-703-55

Query Match 100.0%; Score 81; DB 13; Length 18;
 Best Local Similarity 100.0%; Pred. No. 3e-05;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSGAGIAQAAAGTVNI 17
 |||||
 Db 1 TGSGAGIAQAAAGTVNI 17

RESULT 5
 US-09-287-849-6
 ; Sequence 6, Application US/09287849
 ; Patent No. US20020009459A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 ; FILE REFERENCE: 014058-009020US
 ; CURRENT APPLICATION NUMBER: US/09/287,849
 ; CURRENT FILING DATE: 1999-04-07
 ; PRIOR APPLICATION NUMBER: US 08/818,112
 ; PRIOR FILING DATE: 1997-03-13
 ; PRIOR APPLICATION NUMBER: US 08/942,578
 ; PRIOR FILING DATE: 1997-10-01
 ; PRIOR APPLICATION NUMBER: US 09/025,197
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 09/056,556
 ; PRIOR FILING DATE: 1998-04-07
 ; PRIOR APPLICATION NUMBER: US 09/223,040
 ; PRIOR FILING DATE: 1998-12-30
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 374
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
 US-09-287-849-6

Query Match 100.0%; Score 81; DB 9; Length 374;
 Best Local Similarity 100.0%; Pred. No. 0.00068;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSGAGIAQAAAGTVNI 17
 |||||
 Db 86 TGSGAGIAQAAAGTVNI 102

RESULT 6
 US-09-287-849-40
 ; Sequence 40, Application US/09287849
 ; Patent No. US20020009459A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 ; FILE REFERENCE: 014058-009020US
 ; CURRENT APPLICATION NUMBER: US/09/287,849
 ; CURRENT FILING DATE: 1999-04-07
 ; PRIOR APPLICATION NUMBER: US 08/818,112
 ; PRIOR FILING DATE: 1997-03-13
 ; PRIOR APPLICATION NUMBER: US 08/942,578
 ; PRIOR FILING DATE: 1997-10-01
 ; PRIOR APPLICATION NUMBER: US 09/025,197
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 09/056,556

;; PRIOR FILING DATE: 1998-04-07
;; PRIOR APPLICATION NUMBER: US 09/223,040
;; PRIOR FILING DATE: 1998-12-30
;; NUMBER OF SEQ ID NOS: 46
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 40
;; LENGTH: 374
;; TYPE: PRT
;; ORGANISM: Mycobacterium tuberculosis
;; FEATURE:
;; OTHER INFORMATION: 38 kD antigen
US-09-287-849-40

Query Match 100.0%; Score 81; DB 9; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.0068;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGIAQAAGTUNI 17
| | | | | | | | | | | | | | | | | | | | | |
DB 86 TGSAGIAQAAGTUNI 102

RESULT 7

US-10-193-002-148
; Sequence 148, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, David C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedwick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION NUMBER:
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELEPHONE: (206) 682-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 148:
US-10-193-002-148

Query Match 100.0%; Score 81; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.0068;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGIAQAAGTUNI 17
| | | | | | | | | | | | | | | | | | | | | |
DB 86 TGSAGIAQAAGTUNI 102

RESULT 8

US-10-193-002-150
; Sequence 150, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, David C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedwick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-8031
; INFORMATION FOR SEQ ID NO: 150:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 150:
US-10-193-002-150

Query Match 100.0%; Score 81; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.0068;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGIAQAAGTUNI 17
| | | | | | | | | | | | | | | | | | | | | |
DB 86 TGSAGIAQAAGTUNI 102

RESULT 9

US-10-084-843-153

Sequence 153, Application US/10084843
Publication No. US20030143243A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 153:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 153:
US-10-084-843-153
Query Match 100.0%; Score 81; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSGGAGIAQAAAGTVNI 17
DB 86 TSGGAGIAQAAAGTVNI 102
RESULT 10
US-10-084-843-155
Sequence 155, Application US/10084843
Publication No. US20030143243A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 155:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 155:
US-10-084-843-155
Query Match 100.0%; Score 81; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSGGAGIAQAAAGTVNI 17
DB 86 TSGGAGIAQAAAGTVNI 102
RESULT 11
US-10-359-460-6
Sequence 6, Application US/10359460
Publication No. US20030147911A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/10/359,460
CURRENT FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: US/09/287,849
PRIOR FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07

; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-460-6

Query Match 100.0%; Score 81; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGIAQAAGTNI 17
Db 86 TGSAGIAQAAGTNI 102

RESULT 12
US-10-359-460-40
; Sequence 40, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; PRIOR FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 40
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: 38 kD antigen
US-10-359-460-40

Query Match 100.0%; Score 81; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGIAQAAGTNI 17
Db 86 TGSAGIAQAAGTNI 102

RESULT 13
US-10-098-732A-39
; Sequence 39, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 39
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: 38 kD
US-10-098-732A-39

Query Match 100.0%; Score 81; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGIAQAAGTNI 17
Db 86 TGSAGIAQAAGTNI 102

RESULT 14
US-10-193-002-350
; Sequence 350, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: UNKNOWNH>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 350:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 350:
US-10-193-002-350

Query Match 100.0%; Score 81; DB 14; Length 652;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGAGIAQAAAGTVNI 17
DB 70 TGSAGAGIAQAAAGTVNI 86

RESULT 15
US-10-084-843-355
Sequence 355, Application US/10084843
Publication No. US20030143243A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 355:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 355:
US-10-084-843-355

Query Match 100.0%; Score 81; DB 14; Length 652;

Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGSAGAGIAQAAAGTVNI 17
DB 70 TGSAGAGIAQAAAGTVNI 86

Search completed: March 10, 2004, 12:41:31
Job time : 22.359 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:50:51 ; Search time 40.4295 Seconds
(without alignments)
118.807 Million cell updates/sec

Title: US-10-044-703-49

Perfect score: 81

Sequence: 1 TGSAGIAQAACTVNI 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81	100.0	17	4	AAE12264
2	81	100.0	18	4	AAE12270
3	81	100.0	351	6	ABUG3550
4	81	100.0	373	6	ABUG3549
5	81	100.0	374	2	AAR30090
6	81	100.0	374	2	AAR30090
7	81	100.0	374	2	AAW64363
8	81	100.0	374	2	AAW64364
9	81	100.0	374	2	AAW81730
10	81	100.0	374	2	AAW81731
11	81	100.0	374	2	AAW81731
12	81	100.0	374	2	AAW81731
13	81	100.0	374	2	AAW81731
14	81	100.0	374	2	AAW81731
15	81	100.0	374	2	AAW81731
16	81	100.0	374	5	AAE11840
17	81	100.0	374	5	AAE11840
18	81	100.0	374	5	AAE11840
19	81	100.0	374	5	AAE11840
20	81	100.0	652	2	AAW39082
21	81	100.0	652	2	AAW39082
22	81	100.0	802	2	AAW64379
23	81	100.0	802	2	AAW64379
24	81	100.0	802	2	AAW64379
25	81	100.0	802	2	AAW64379

26	81	100.0	802	2	AAW39081
27	81	100.0	802	2	AAW39176
28	81	100.0	802	2	AAW39224
29	81	100.0	802	5	AAU74592
30	81	100.0	983	4	AAU01901
31	51	63.0	283	5	ABW55069
32	47	58.0	298	5	ABW55070
33	47	58.0	301	7	ADC95326
34	47	58.0	1407	6	ABU19785
35	46	56.8	239	4	AAW90998
36	46	56.8	435	4	ABW60210
37	46	56.8	966	7	AAE73004
38	46	56.8	3010	2	AAW30616
39	45	55.6	261	2	AAW00079
40	45	55.6	261	5	ABP43298
41	45	55.6	261	6	ABU88326
42	45	55.6	261	6	ABU13577
43	45	55.6	284	2	AAW00078
44	45	55.6	284	5	ABP43297
45	45	55.6	284	6	ABU88325

ALIGNMENTS

RESULT 1

AAE12264
ID AAE12264 standard; peptide; 17 AA.

XX AAE12264;

AC AAE12264;

DT 18-DEC-2001 (first entry)

XX AAE12264;

DE Mycobacterium tuberculosis (Mtb) peptide #49.

XX AAE12264;

KW Mycobacterium tuberculosis; Mtb peptide; antibacterial; vaccine; infection; anti-Mtb immune response.

XX AAE12264;

OS Mycobacterium tuberculosis.

PN WO200170774-A2.

XX AAE12264;

PD 27-SEP-2001.

XX AAE12264;

PF 20-MAR-2001; 2001WO-US008906.

XX AAE12264;

PR 20-MAR-2000; 2000US-0190834P.

XX AAE12264;

PA (UYER-) UNIV BROWN RES FOUND.

XX AAE12264;

PI Degroot AS;

XX AAE12264;

DR WPI; 2001-616401/71.

XX AAE12264;

PT New vaccine for immunising a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis.

XX AAE12264;

PS Disclosure; Fig 4; 42pp; English.

XX AAE12264;

CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mycobacterium tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

XX AAE12264;

CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mycobacterium tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

XX AAE12264;

CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mycobacterium tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

XX AAE12264;

CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mycobacterium tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

XX AAE12264;

CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mycobacterium tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

XX AAE12264;

CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mycobacterium tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

XX AAE12264;

CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mycobacterium tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

XX AAE12264;

CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mycobacterium tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

XX AAE12264;

CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mycobacterium tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

XX AAE12264;

CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mycobacterium tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

XX AAE12264;

CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mycobacterium tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

XX AAE12264;

CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mycobacterium tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

Query Match 100.0%; Score 81; DB 4; Length 17;

Best Local Similarity 100.0%; Pred. No. 8.2e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGAGIAQAAAGTVNI 17
DB 1 TGSAGAGIAQAAAGTVNI 17

RESULT 2
AAE12270
ID AAE12270 standard; peptide; 18 AA.
XX AC AAE12270;
DT 18-DEC-2001 (first entry)
XX XX
DE Mycobacterium tuberculosis (Mtb) peptide #55.
XX XX
KW Mycobacterium tuberculosis; Mtb peptide; antibacterial; vaccine;
XX infection; anti-Mtb immune response.
XX OS Mycobacterium tuberculosis.
XX XX
PN WO200170774-A2.
XX PD 27-SEP-2001.
XX XX
PF 20-MAR-2001; 2001WO-US008906.
XX XX
PR 20-MAR-2000; 2000US-0190834P.
XX PA (UYER-) UNIV BROWN RES FOUND.
XX PI Degroot AS;
XX DR WPI; 2001-616401/71.
XX XX
PT New vaccine for immunizing a mammalian subject, preferably humans,
PT against infection caused by Mycobacterium tuberculosis.
XX PS Disclosure; Fig 4; 42pp; English.
XX CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine
CC candidate peptides. The invention also relates to a method for
CC identifying Mtb vaccine candidate peptides as well as vaccines comprising
CC these candidate peptides. Vaccines of the invention and Mtb vaccine
CC candidate peptides are useful for inducing an anti- Mycobacterium
CC tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a
CC mammalian subject preferably human. They are used for immunising a
CC mammalian subject, preferably humans, against infection caused by
CC Mycobacterium tuberculosis. The present sequence is a Mtb vaccine
CC candidate peptide
XX XX
SQ Sequence 18 AA;

Query Match 100.0%; Score 81; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 8.7e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGAGIAQAAAGTVNI 17
DB 1 TGSAGAGIAQAAAGTVNI 17

RESULT 3
ABU63550
ID ABU63550 standard; protein; 351 AA.
XX AC ABU63550;
XX XX
DT 25-SEP-2003 (first entry)
XX XX
DE Mature Mycobacterium tuberculosis 38kDa protein.

XX Antimicrobial; cytostatic; type 1 response; 38kDa protein;
KW T-helper-1 T-cell response; interleukin-12; IL-12;
KW cell-mediated immunity; pathogen infection; vaccine;
KW interferon-gamma sensitive tumour.
XX XX
OS Mycobacterium tuberculosis.
XX XX
FH Key Location/Qualifiers
FT Modified-site 1
FT /label= OTHER
FT /note= "OTHER= N-acyl diglyceride cysteine"
XX XX
PN US6517839-B1.
XX XX
PD 11-FEB-2003.
XX XX
PF 17-JUL-1998; 98US-00118426.
XX XX
PR 19-JUL-1997; 97US-0052970P.
XX XX
PA (REGC) UNIV CALIFORNIA.
XX XX
PI Modlin RL, Libraty DH;
XX DR WPI; 2003-553609/52.
XX XX
PT Inducing interleukin-12 and type 1/T-helper-1 T-cell response to
PT stimulate cell-mediated immunity for preventing and treating pathogen
PT infection involves administering lipopeptide having N-terminal ester- or
PT amide-linked fatty acyl group.
XX XX
PS Disclosure; Col 25-26; 27pp; English.
XX XX
CC The invention describes a method of inducing type 1/T-helper-1 T-cell
CC response or interleukin-12 (IL-12) in a cell involves administration of a
CC lipopeptide (1) having an N-terminal ester- or amide-linked fatty acyl
CC group. The methods are used to stimulate cell-mediated immunity for
CC prevention and treatment of pathogen infections (e.g. Mycobacterium
CC tuberculosis) and for treatment of a interferon-gamma sensitive tumour.
CC Also useful in the development of new vaccine and therapeutic strategies.
CC This is the amino acid sequence of mature Mycobacterium tuberculosis
CC 38kDa protein from which lipopeptides can be isolated
XX XX
SQ Sequence 351 AA;
Query Match 100.0%; Score 81; DB 6; Length 351;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGAGIAQAAAGTVNI 17
DB 63 TGSAGAGIAQAAAGTVNI 79

RESULT 4
ABU63549
ID ABU63549 standard; protein; 373 AA.
XX AC ABU63549;
XX XX
DT 25-SEP-2003 (first entry)
XX XX
DE Mycobacterium tuberculosis 38kDa protein.
XX XX
KW Antimicrobial; cytostatic; type 1 response; 38kDa protein;
KW T-helper-1 T-cell response; interleukin-12; IL-12;
KW cell-mediated immunity; pathogen infection; vaccine;
KW interferon-gamma sensitive tumour.
XX XX
OS Mycobacterium tuberculosis.
XX XX
FH Key Location/Qualifiers

FT Modified-site 1 /label= OTHER
 FT /note= "OTHER= N-acyl diglyceride cysteine"
 PN US6517839-B1.
 XX 11-FEB-2003.
 XX 17-JUL-1998; 98US-00118426.
 XX 18-JUL-1997; 97US-0052970P.
 PR (REGC) UNIV CALIFORNIA.
 PA Modlin RL, Libraty DH;
 XX WPI; 2003-553609/52.
 DR
 XX Inducing interleukin-12 and type 1/T-helper-1 T-cell response to
 FT stimulate cell-mediated immunity for preventing and treating pathogen
 PT infection involves administering lipopeptide having N-terminal ester- or
 PT amide-linked fatty acyl group.
 XX Disclosure; Col 23-24; 27pp; English.
 XX
 CC The invention describes a method of inducing type 1/T-helper-1 T-cell
 CC response or interleukin-12 (IL-12) in a cell involves administration of a
 CC lipopeptide (I) having an N-terminal ester- or amide-linked fatty acyl
 CC group. The methods are used to stimulate cell-mediated immunity for
 CC prevention and treatment of pathogen infections (e.g. Mycobacterium
 CC tuberculosis) and for treatment of a interferon-gamma sensitive tumour.
 CC Also useful in the development of new vaccine and therapeutic strategies.
 CC This is the amino acid sequence of Mycobacterium tuberculosis 38kDa
 CC protein from which lipopeptides can be isolated
 XX
 SQ Sequence 373 AA;
 Query Match 100.0%; Score 81; DB 6; Length 373;
 Best Local Similarity 100.0%; Pred. No. 0.0017;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 TSGGAGIAQAAGTVNI 17
 DB 85 TSGGAGIAQAAGTVNI 101
 RESULT 5
 AAR30090
 ID AAR30090 standard; protein; 374 AA.
 XX AC AAR30090;
 XX 25-MAR-2003 (revised)
 DT 05-MAY-1993 (first entry)
 XX 38 kDa lipoprotein antigen.
 DE
 XX Tubercule bacilli protein; tuberculosis; diagnosis; TB; peptide.
 KW
 XX Mycobacterium tuberculosis.
 OS
 XX Key Location/Qualifiers
 FH Peptide 1..20 /note= "peptide A"
 FT Peptide 24..43 /note= "peptide H"
 FT Region 66..76 /note= "amphipathic region"
 FT Peptide 146..165 /note= "peptide B"
 FT Region 150..164 /note= "amphipathic region"
 FT Peptide 201..220

FT Region /note= "peptide C"
 FT 204..210 /note= "amphipathic region"
 FT Peptide 210..229 /note= "peptide D"
 FT Region 216..225 /note= "amphipathic region"
 FT Peptide 230..249 /note= "peptide E"
 FT Region 236..243 /note= "amphipathic region"
 FT Peptide 285..304 /note= "peptide F"
 FT Region 287..291 /note= "amphipathic region"
 FT Region 296..304 /note= "amphipathic region"
 FT Peptide 350..369 /note= "peptide G"
 FT Region 362..368 /note= "amphipathic region"
 XX
 PN WO9221697-A2.
 XX
 XX 10-DEC-1992.
 XX 26-MAY-1992; 92WO-GB000948.
 PF 24-MAY-1991; 91GB-00011291.
 PR (MEDI-) MEDICAL RES COUNCIL.
 XX Vordermeier H, Harris D, Moreno C, Ivanyi J;
 PI WPI; 1992-433610/52.
 XX
 XX Peptide(s) from mycobacterium tuberculosis antigens - useful for the
 PT diagnosis of tuberculosis.
 XX Disclosure; Fig 1; 44pp; English.
 XX
 CC The sequence is that of the 38 kDa lipoprotein antigen of Mycobacterium
 CC tuberculosis, from this peptide can be derived which can be used in
 CC place of the purified protein derivative (PPD) test. The peptides can be
 CC used to diagnose tuberculosis (TB) in a human or non-human animal, and to
 CC distinguish a TB patient from an infected or otherwise sensitised but
 CC healthy clinical suspect. They stimulate lymphocytes which have been
 CC sensitised to antigens of M. tuberculosis. The peptides may be used in
 CC vivo skin tests relying on delayed hypersensitivity reaction causing
 CC reddening and swelling, or in ex vivo detection of activated lymphocytes.
 CC They have excellent sensitivity and improved specificity to PPD-
 CC tuberculin. They fail to provoke a strong immune response in delayed-type
 CC hypersensitivity or lymphocyte activation tests in patients with
 CC pulmonary and non-lymphatic extrapulmonary TB. This provides the basis of
 CC distinguishing TB patients from these patients. Diagnosing TB comprises
 CC either intradermal injection of the peptide, or contacting lymphocytes
 CC with the peptide. (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 81; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 0.0017;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 TSGGAGIAQAAGTVNI 17
 DB 86 TSGGAGIAQAAGTVNI 102
 RESULT 6
 AAW64363
 ID AAW64363 standard; protein; 374 AA.
 XX

AAW64363;
 09-NOV-1998 (first entry)
 Mycobacterium antigen TbrA3-38 kD-Tb38-1 fusion protein.
 Tuberculosis; infection; diagnosis; 38 kDa antigen; TbrA3; Tb38-1.
 Mycobacterium tuberculosis.
 Synthetic.
 WO9816645-A2.
 23-APR-1998.
 07-OCT-1997; 97WO-US018214.
 11-OCT-1996; 96US-00729622.
 13-MAR-1997; 97US-00818111.
 (CORI-) CORIXA CORP.
 Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 Vedwick TS, Twardzik DR, Lodes MJ;
 WPI; 1998-251292/22.
 N-PSDB; AAV44413.
 New isolated Mycobacterium tuberculosis polypeptides and DNA - used to develop products for the detection of M. tuberculosis infection and diagnosis of tuberculosis.
 Example 7; Page 165-166; 250pp; English.
 This polypeptide comprises a fusion protein composed of Mycobacterium tuberculosis antigens TbrA3 (see AAW64295), 38 kDa antigen (see AAW64364) and Tb38-1 (see AAW64321). It was produced by PCR amplification (see AAV44407-12) of the appropriate antigen DNA sequences (see AAV44413) and cloning into an expression vector. The fusion protein was expressed in E. coli. It can be used for serodiagnosis of tuberculosis. The invention relates to compositions and methods for diagnosing tuberculosis. It provides polypeptides (see AAW64291-W64379) comprising an antigenic portion of a soluble M. tuberculosis antigen, an immunogenic portion of an M. tuberculosis antigen, or a fusion protein, as well as DNA sequences encoding such polypeptides, recombinant expression vectors and transformed or transfected host cells. Also claimed are methods and diagnostic kits for detecting M. tuberculosis infection in a patient

Query Match 100.0%; Score 81; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 0.0017;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSGAGIAQAAAGTVNI 17
 |||||
 Db 86 TSGAGIAQAAAGTVNI 102

RESULT 7
 AAW64364
 ID AAW64364 standard; protein; 374 AA.
 AC AAW64364;
 XX
 XX
 DT 27-JAN-1999 (first entry)
 XX
 DE M. tuberculosis fusion protein TbrA3/38kD/Tb38-1.
 XX
 KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 KW vaccine; pharmaceutical; infection; diagnosis.
 XX
 OS Synthetic.
 OS Mycobacterium tuberculosis.
 XX
 FN WO9816646-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US018293.
 XX
 PR 11-OCT-1996; 96US-00730510.
 PR 13-MAR-1997; 97US-00818112.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

PN WO9816645-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US018214.
 XX
 PR 11-OCT-1996; 96US-00729622.
 PR 13-MAR-1997; 97US-00818111.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedwick TS, Twardzik DR, Lodes MJ;
 XX
 DR 1998-251292/22.
 DR N-PSDB; AAV44414.
 XX
 PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used to develop products for the detection of M. tuberculosis infection and diagnosis of tuberculosis.
 PT
 PS Disclosure; Page 168-169; 250pp; English.
 XX
 CC This polypeptide comprises Mycobacterium tuberculosis 38 kDa antigen. A TbrA3-38 kD-Tb38-1 fusion protein (see AAW64363) has been constructed that can be used in the serodiagnosis of tuberculosis. The invention relates to compositions and methods for diagnosing tuberculosis. It provides polypeptides (see AAW64291-W64379) comprising an antigenic portion of a soluble M. tuberculosis antigen, an immunogenic portion of an M. tuberculosis antigen, or a fusion protein, as well as DNA sequences encoding such polypeptides, recombinant expression vectors and transformed or transfected host cells. Also claimed are methods and diagnostic kits for detecting M. tuberculosis infection in a patient

Query Match 100.0%; Score 81; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 0.0017;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSGAGIAQAAAGTVNI 17
 |||||
 Db 86 TSGAGIAQAAAGTVNI 102

RESULT 8
 AAW81730
 ID AAW81730 standard; protein; 374 AA.
 AC AAW81730;
 XX
 XX
 DT 27-JAN-1999 (first entry)
 XX
 DE M. tuberculosis fusion protein TbrA3/38kD/Tb38-1.
 XX
 KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 KW vaccine; pharmaceutical; infection; diagnosis.
 XX
 OS Synthetic.
 OS Mycobacterium tuberculosis.
 XX
 FN WO9816646-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US018293.
 XX
 PR 11-OCT-1996; 96US-00730510.
 PR 13-MAR-1997; 97US-00818112.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX WPI; 1998-261042/23.
 DR N-PSDB; AAV64522.
 XX
 PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and for
 PT diagnosis, treatment and prevention of tuberculosis.
 XX
 PS Claim 37; Page 156-157; 230pp; English.
 XX
 CC This sequence represents a fusion protein containing the immunogenic
 CC polypeptides TBRa3, 38kD and Tb38-1 from Mycobacterium tuberculosis (MT).
 CC This fusion protein can be used in a method for inducing protective
 CC immunity against tuberculosis (TB). This sequence can be formulated into
 CC vaccines and/or pharmaceutical compositions for immunising against M.
 CC tuberculosis infection or may be used for the diagnosis of TB
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 81; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. NO. 0.0017;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGSAGIAQAAGTNI 17
 DB 86 TGSAGIAQAAGTNI 102
 RESULT 9
 AAWS1731
 ID AAWS1731 standard; protein; 374 AA.
 XX
 AC AAWS1731;
 XX
 DT 27-JAN-1999 (first entry)
 XX
 DE M. tuberculosis 38kD antigen protein.
 XX
 KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 KW vaccine; pharmaceutical; infection; diagnosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9816646-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US018293.
 XX
 PR 11-OCT-1996; 96US-00730510.
 PR 13-MAR-1997; 97US-00818112.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAM, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX
 DR WPI; 1998-261042/23.
 DR N-PSDB; AAV64523.
 XX
 CC Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
 CC develop products for the detection of M. tuberculosis infection and for
 CC diagnosis, treatment and prevention of tuberculosis.
 XX
 PS Claim 23; Page 159-160; 230pp; English.
 XX
 CC This sequence represents a 38kD antigen from Mycobacterium tuberculosis
 CC (MT). This antigen can be used in a method for inducing protective
 CC immunity against tuberculosis (TB). This sequence can be formulated into
 CC vaccines and/or pharmaceutical compositions for immunising against M.
 CC tuberculosis infection or may be used for the diagnosis of TB
 XX

SQ Sequence 374 AA;
 Query Match 100.0%; Score 81; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. NO. 0.0017;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGSAGIAQAAGTNI 17
 DB 86 TGSAGIAQAAGTNI 102
 RESULT 10
 AAY32061
 ID AAY32061 standard; protein; 374 AA.
 XX
 AC AAY32061;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen TBRa3-38kD-Tb38-1 fusion.
 XX
 KW Tuberculosis; antigen; fusion protein; TBRa3; 38kD; Tb38-1; diagnosis;
 KW therapy; vaccine; immunogen.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9951748-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 07-APR-1999; 99WO-US007717.
 XX
 PR 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky YAM, Alderson M, Campos-Neto A;
 XX
 DR WPI; 1999-601610/51.
 DR N-PSDB; AAZ20196.
 XX
 PT New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis.
 XX
 PS Claim 1; Fig 3A-D; 83pp; English.
 XX
 CC This sequence represents a recombinant Mycobacterium tuberculosis tri-
 CC antigen fusion protein composed of the antigens TBRa3, 39kD and Tb38-1.
 CC The fusion protein is expressed in host cells using a vector carrying a
 CC polynucleotide (see AAZ20196) comprising the 3 coding sequences for the
 CC antigens. The invention provides fusion proteins (see AAY32059-71)
 CC containing at least 2 M. tuberculosis antigens. The new fusion proteins
 CC and polynucleotides encoding them are useful as vaccines for preventing
 CC tuberculosis (claimed), for diagnosis (via in vitro assays or intradermal
 CC skin tests for detection of anti-M. tuberculosis antibodies), monitoring
 CC of disease progression, and treatment of tuberculosis. They are more
 CC effective immunogens than mixtures of the individual protein components
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 81; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. NO. 0.0017;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGSAGIAQAAGTNI 17
 DB 86 TGSAGIAQAAGTNI 102
 RESULT 11
 AAY39018
 ID AAY39018 standard; protein; 374 AA.

XX AAY39018;
 AC
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis 38 kDa antigen protein.
 XX
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 XX vaccine; immunity.
 KW
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942118-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US003265.
 XX
 PR 18-FEB-1998; 98US-00024753.
 XX
 PR 05-MAY-1998; 98US-00072596.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX
 DR WPI; 1999-527416/44.
 DR N-PSDB; AA219112.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis.
 XX
 PS Claim 53; Page 199-200; 323pp; English.
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against M.
 CC tuberculosis infection. The new detection methods are needed as current
 CC vaccination strategies do not provide 100% immunity
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 81; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 0.0017;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGSAGAGIAQAAGTNYI 17
 DB 86 TGSAGAGIAQAAGTNYI 102
 RESULT 12
 AAY39017
 ID AAY39017 standard; protein; 374 AA.
 AC AAY39017;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis fusion protein TBra3/38kDa/Tb38-1.
 XX
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 KW
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942118-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US003265.
 XX

PR 18-FEB-1998; 98US-00024753.
 PR 05-MAY-1998; 98US-00072596.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX
 DR WPI; 1999-527416/44.
 DR N-PSDB; AA219111.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis.
 XX
 PS Example 10; Page 196-198; 323pp; English.
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against M.
 CC tuberculosis infection. The new detection methods are needed as current
 CC vaccination strategies do not provide 100% immunity
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 81; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 0.0017;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGSAGAGIAQAAGTNYI 17
 DB 86 TGSAGAGIAQAAGTNYI 102
 RESULT 13
 AAY39161
 ID AAY39161 standard; protein; 374 AA.
 XX
 AC AAY39161;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis antigen 38 kD amino acid sequence.
 XX
 KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942076-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US003268.
 XX
 PR 18-FEB-1998; 98US-00025197.
 PR 05-MAY-1998; 98US-00072967.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX
 DR WPI; 1999-527409/44.
 DR N-PSDB; AA219324.
 XX
 PT New antigens from Mycobacterium tuberculosis useful in diagnostic skin
 PT tests and protective or therapeutic vaccines or compositions.
 XX
 PS Claim 23; Page 154-155; 299pp; English.
 XX
 CC The present invention describes polypeptides comprising an immunogenic

CC Part of a Mycobacterium tuberculosis antigen (Ag). Also described are
 CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
 CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
 CC polypeptides fragments, can be used in pharmaceutical compositions or
 CC vaccines to generate a protective or therapeutic immune response to M.
 CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
 CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
 CC killer cells and/or macrophages in tuberculosis-immune subjects. AA219249
 CC to AA219460 and AA219083 to AA219225 are used in the exemplification of
 CC the present invention

XX SQ Sequence 374 AA;

Query Match 100.0%; Score 81; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 0.0017;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGAGIAQAAAGTNNI 17
 |||||
 Db 86 TGSAGAGIAQAAAGTNNI 102

RESULT 14
 AAY39160
 ID AAY39160 standard; protein; 374 AA.

XX AC AAY39160;

XX DT 05-NOV-1999 (first entry)

XX DE M. tuberculosis Tbra3, 38 kD and Tbra3-1 fusion protein.

XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 XX KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 XX KW immune response; skin test.

XX OS Synthetic.

XX OS Mycobacterium tuberculosis.

XX FN WO9942076-A2.

XX PD 26-AUG-1999.

XX PF 17-FEB-1999; 99WO-US003268.

XX PR 18-FEB-1998; 98US-00025197.

XX PR 05-MAY-1998; 98US-00072967.

XX PA (CORI-) CORIXA CORP.

XX PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;

XX DR WPI; 1999-527409/44.

XX DR N-PSDB; AA219323.

XX PT New antigens from Mycobacterium tuberculosis useful in diagnostic skin
 XX tests and protective or therapeutic vaccines or compositions.

XX PS Claim 37; Page 151-152; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
 CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
 CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
 CC polypeptides fragments, can be used in pharmaceutical compositions or
 CC vaccines to generate a protective or therapeutic immune response to M.
 CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
 CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
 CC killer cells and/or macrophages in tuberculosis-immune subjects. AA219249
 CC to AA219460 and AA219083 to AA219225 are used in the exemplification of
 CC the present invention

XX

SQ Sequence 374 AA;

Query Match 100.0%; Score 81; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 0.0017;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGAGIAQAAAGTNNI 17
 |||||
 Db 86 TGSAGAGIAQAAAGTNNI 102

RESULT 15

AAE11840

ID AAE11840 standard; protein; 374 AA.

XX AC AAE11840;

XX DT 18-DEC-2001 (first entry)

XX DE Mycobacterium tuberculosis 38kDa-LP protein.

XX KW 38kDa-LP protein; antibiotic; vaccine; CD8 T-cell; epitope; medicament;
 XX KW prophylactic; therapeutic; mycobacterial infection.

XX OS Mycobacterium tuberculosis.

XX FN WO200170764-A2.

XX PD 27-SEP-2001.

XX PF 20-MAR-2001; 2001WO-GB001205.

XX PR 20-MAR-2000; 2000GB-00006692.

XX PA (GLAX) GLAXO GROUP LTD.

XX PI Klein MR;

XX DR WPI; 2001-607517/69.

XX DR N-PSDB; AAD18885.

XX PT Use of polypeptide comprising CD8 T-cell epitopes derived from
 XX Mycobacterium tuberculosis proteins in the manufacture of a medicament
 XX for vaccinating prophylactically or therapeutically against mycobacterial
 XX infection.

XX PS Claim 3; Page 50-52; 54pp; English.

XX The invention relates to use of polypeptides comprising CD8 T-cell
 CC epitopes derived from Mycobacterium tuberculosis proteins. The invention
 CC is useful in the manufacture of a medicament for prophylactic or
 CC therapeutic vaccination against mycobacterial infection. The polypeptide
 CC and the expression vector are capable of stimulating CD8 T-cell response.
 CC The invention also provides a vaccine composition which comprises
 CC polypeptide or expression vector useful for vaccinating a pre-selected
 CC host against a mycobacterial infection. The invention further relates to
 CC a method for determining the presence or absence of CD8 T-cell response
 CC to epitope sequence by identifying the presence or absence of T-cell that
 CC recognise the epitope sequence in a sample from the host. The present
 CC sequence is Mycobacterium tuberculosis 38kDa-LP protein related to the
 CC invention

XX SQ Sequence 374 AA;

Query Match 100.0%; Score 81; DB 4; Length 374;
 Best Local Similarity 100.0%; Pred. No. 0.0017;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGAGIAQAAAGTNNI 17
 |||||
 Db 86 TGSAGAGIAQAAAGTNNI 102

Search completed: March 10, 2004, 12:05:30
Job time : 41.4295 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:58:01 ; Search time 9.15385 Seconds
(without alignments)
178.641 Million cell updates/sec

Title: US-10-044-703-49

Perfect score: 81

Sequence: 1 TGGAGTAAACTVNI 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81	100.0	374	1 F70584	phosphate specific
2	51	63.0	283	2 H86839	hypothetical prote
3	49	60.5	347	2 AD1920	ABC phosphate tran
4	47	58.0	298	2 A86840	hypothetical prote
5	47	58.0	806	2 B49938	hupU protein - Rho
6	46	56.8	364	2 G82595	ABC transporter ph
7	45	55.6	292	2 A95163	hypothetical prote
8	45	55.6	292	2 H98028	hypothetical prote
9	45	55.6	304	2 AH3302	7alpha-hydroxyster
10	45	55.6	392	2 AG2377	phosphate-binding
11	44	54.3	300	1 A69956	phosphate ABC tran
12	44	54.3	640	2 T08179	LKGS protein - Chl
13	44	54.3	708	1 JC2465	copper-transportin
14	44	54.3	718	2 A36068	major ampullate fi
15	44	54.3	854	2 D82324	conserved hypothet
16	44	54.3	1408	2 T45039	hypothetical prote
17	43	53.1	177	2 D83015	heat shock protein
18	43	53.1	364	2 A81019	adhesin/invasin, p
19	43	53.1	471	2 AE3384	zinc metalloprote
20	42	51.9	254	2 D69648	2-keto-3-deoxygluc
21	42	51.9	261	2 AD2966	dehydrogenase Atu3
22	42	51.9	261	2 H98316	hypothetical prote
23	42	51.9	307	1 DEPGC	3-hydroxyacyl-CoA
24	42	51.9	314	1 JC4879	3-hydroxyacyl-CoA
25	42	51.9	314	1 JC4210	3-hydroxyacyl-CoA
26	42	51.9	314	2 T46866	3-hydroxyacyl-CoA
27	42	51.9	414	2 AG2996	transcription regu
28	42	51.9	429	2 C87183	phosphate-binding
29	42	51.9	467	2 B95850	probable amino aci

30 42 51.9 467 2 B98287 probable transcript
31 42 51.9 488 2 D35385 hydrogen dehydroge
32 42 51.9 526 2 AE3004 conserved hypothet
33 42 51.9 526 2 C98279 hypothetical prote
34 42 51.9 788 2 AC2767 formate dehydrogen
35 42 51.9 788 2 F97547 probable oxidoredu
36 42 51.9 834 2 S19155 serotonin receptor
37 42 51.9 1174 2 A40853 potassium channel
38 42 51.9 1819 2 T26533 hypothetical glyci
39 42 51.9 2174 2 B95965 hypothetical glyci
40 42 51.9 3010 1 GNVVTC genome polyprotein
41 42 51.9 3010 1 GNVVTC genome polyprotein
42 42 51.9 3010 1 S18030 genome polyprotein
43 42 51.9 3010 1 GNVVTC genome polyprotein
44 41 50.6 112 2 F87193 probable secreted
45 41 50.6 146 1 S24311 flavodoxin - Desul

ALIGNMENTS

RESULT 1

F70584
phosphate specific transporter S precursor - Mycobacterium tuberculosis (strain H37RV)
N.Alternate names: antigen b
C.Species: Mycobacterium tuberculosis
C.Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 16-Jun-2000
C.Accession: F70584; JC5103; A42930; A49721; A45820
R.Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, C.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A.Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A.Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome :
A.Reference number: A70500; MUID:98295987; PMID:9634230
A.Accession: F70584
A>Status: nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-374 <COL>
A.Cross-references: GB:Z95209; GB:AL123456; NID:G3261750; PIDN:CA808484.1; PID:G2078049
A.Experimental source: strain H37RV
R.Braibant, M.; Lefevre, P.; de Wit, L.; Peirs, P.; Ooms, J.; Huygen, K.; Andersen, A.B.;
Gene 176, 171-176, 1996
A.Title: A Mycobacterium tuberculosis gene cluster encoding proteins of a phosphate trans
A.Reference number: JC5100; MUID:97075926; PMID:8918249
A.Accession: JC5103
A.Molecule type: DNA
A.Residues: 1-374 <BRA>
A.Cross-references: GB:M30046; NID:G149987; PIDN:AAA25374.1; PID:G149988
A.Note: neither the complete nucleic acid sequence nor the complete translation are show
A.Note: the authors translated the initiation codon GTG for residue 1 as Met
R.Andersen, A.B.; Hansen, E.B.
Infect. Immun. 57, 2481-2488, 1989
A.Title: Structure and mapping of antigenic domains of protein antigen b, a 38,000-molecu
A.Reference number: A42930; MUID:89307568; PMID:12545626
A.Accession: A42930
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-374 <AND1>
A.Cross-references: GB:M30046; NID:G149987; PIDN:AAA25374.1; PID:G149988
R.Chang, Z.; Choudhary, A.; Lathigra, R.; Quijcho, F.A.
J. Biol. Chem. 269, 1956-1958, 1994
A.Title: The immunodominant 38-kDa lipoprotein antigen of Mycobacterium tuberculosis is :
A.Reference number: A49721; MUID:94124544; PMID:8294447
A.Accession: A49721
A.Molecule type: protein
A.Residues: 25-34 <CHA>
R.Andersen, A.B.; Ljungqvist, L.; Olsen, M.
J. Gen. Microbiol. 136, 477-480, 1990
A.Title: Evidence that protein antigen b of Mycobacterium tuberculosis is involved in phc
A.Reference number: A45820; MUID:90362031; PMID:2118164
A.Accession: A45820
A.Molecule type: protein

A:Residues: 'X',27-28,'X',30-34,'XX',37 <AND2>
A>Note: confirmed presence of normal signal and absence of lipoprotein attachment

C:Genetics:
A:Gene: phoS1; psts
A:Start codon: GTG
C:Superfamily: phosphate-repressible phosphate-binding protein
C:Keywords: surface antigen
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-374/Product: phosphate specific transporter S #status experimental <NAT>

Query Match 100.0%; Score 81; DB 1; Length 374;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGIAQAAGTNNI 17
DB 86 TGSAGIAQAAGTNNI 102

RESULT 2

H86839
Hypothetical protein pstE [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: H86839
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: H86839
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-283 <STO>
A:Cross-references: GB:AE005176; PID:g12724738; PIDN:AAK05818.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:

C:Superfamily: Methanobacterium phosphate-binding protein pstS
Query Match 63.0%; Score 51; DB 2; Length 283;
Best Local Similarity 62.5%; Pred. No. 3.1;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 GSGAGIAQAAGTNNI 17
DB 62 GSGVGLAQVAGSFQI 77

RESULT 3

AD1920
ABC phosphate transport system phosphate-binding periplasmic protein all0911 [imported]

C:Species: Nostoc sp. PCC 7120
A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AD1920
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shampo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD1920
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-347 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA072868.1; PID:g17130257; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0911

C:Superfamily: phosphate-repressible phosphate-binding protein
Query Match 60.5%; Score 49; DB 2; Length 347;
Best Local Similarity 66.7%; Pred. No. 7.4;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 GSGAGIAQAAGTNNI 16
DB 70 GSGGIRQTIACTVD 84

RESULT 4

A86840
Hypothetical protein pstF [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 14-Sep-2001
C:Accession: A86840
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: A86840
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-298 <STO>

A:Cross-references: GB:AE005176; PID:g12724739; PIDN:AAK05819.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: pstF
C:Superfamily: Methanobacterium phosphate-binding protein pstS

Query Match 58.0%; Score 47; DB 2; Length 298;
Best Local Similarity 56.2%; Pred. No. 13;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 GSGAGIAQAAGTNNI 17
DB 77 GSGVGLTQVAGSFQI 92

RESULT 5

B49938
hupU protein - Rhodobacter capsulatus

C:Species: Rhodobacter capsulatus
C:Date: 19-May-1995 #sequence_revision 19-May-1995 #text_change 03-Aug-1995
C:Accession: B49938; S24787
R:Eisen, S.; Richaud, P.; Colbeau, A.; Vignais, P.M.
J. Bacteriol. 175, 7404-7412, 1993
A:Title: Sequence analysis and interposon mutagenesis of the hupU gene, which encodes a
A:Reference number: A49938; MUID:94042916; PMID:8226687
A:Accession: B49938
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-806 <ELS>

A:Cross-references: GB:L02348
A>Note: authors translated the codon CTG for residue 30 as Met, TGC for residue 261 as Trp
R:Richaud, P.
submitted to the EMBL Data Library, January 1991
A:Reference number: S24786
A:Accession: S24787
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-806 <RIC>

A:Cross-references: EMBL:X57380
Query Match 58.0%; Score 47; DB 2; Length 806;
Best Local Similarity 53.3%; Pred. No. 32;
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGSAGIAQAAGTNNI 15
DB 722 TGAGAGLVEAARGSL 736

RESULT 6

G82595
ABC transporter phosphate binding protein XF2141 [imported] - Xylella fastidiosa (strain
C:Species: Xylella fastidiosa

```
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: G82595
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: G82595
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-364 <STM>
A:Cross-references: GB:AE004028; GB:AE003849; NID:g9107266; PIDN:AAF84940.1; GSPDB:GN001
A:Experimental source: strain 3a5C
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Paciniani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.W.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2141
C:Superfamily: phosphate-repressible phosphate-binding protein
Query Match 56.8%; Score 46; DB 2; Length 364;
Best Local Similarity 60.0%; Pred. No. 21;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 2 GSGAGIAQAAGTNN 16
||| ||| ||| ||| |||
DB 86 GSGGIAQIXAATID 100

RESULT 7
A:Title: Hypothetical protein SPI400 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: A95163
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: A95163
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-292 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75498.1; PID:g14972888; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI400
C:Superfamily: Methanobacterium phosphate-binding protein pstS
Query Match 55.6%; Score 45; DB 2; Length 292;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 2 GSGAGIAQAAGTNN 17
||| ||| ||| ||| |||
DB 67 GSGTGLSQVSGAVDI 82

RESULT 8
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
phosphate-binding periplasmic protein of phosphate ABC transporter all4575 [imported] - A
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG2377
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
```


A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AG2377

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-392 <KUR>

A;Cross-references: GB:BA000019; PIDN:BA076274.1; PID:g17133711; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: all4575

C;Superfamily: phosphate-repressible phosphate-binding protein

Query Match 55.6%; Score 45; DB 2; Length 392;

Best Local Similarity 60.0%; Pred. No. 32;

Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 GSGAGTAAAGTNN 16

|||||:|||||

Db 101 GSGAGVEQFIQGTVD 115

RESULT 11

A69956

phosphate ABC transporter (binding protein) homolog yqgG - Bacillus subtilis

C;Species: Bacillus subtilis

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C;Accession: A69956

R;Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter

A.; Bron, S.; Brouiller, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd

Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler

iechter, J.; Haywood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.

Koetter, P.; Koningsstein, G.; Krog, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois

A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel

Y.M.; Osawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portecelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,

A;Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror

akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: A69956

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-300 <KUN>

A;Cross-references: GB:Z99116; GB:AL009126; NID:g2634723; PIDN:CAB14429.1; PID:g2634932

A;Experimental source: strain 168

C;Genetics:

A;Gene: yqgG

C;Superfamily: Methanobacterium phosphate-binding protein pstS

Query Match 54.3%; Score 44; DB 1; Length 300;

Best Local Similarity 50.0%; Pred. No. 35;

Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 GSGAGTAAAGTNN 17

|||||:|||||

Db 80 GSGTGLSQVSEGAVQI 95

RESULT 12

T08179

LRG5 Protein - Chlamydomonas reinhardtii

C;Species: Chlamydomonas reinhardtii

C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999

C;Accession: T08179

R;Glockner, G.; Beck, C.F.

submitted to the EMBL Data Library, October 1996

A;Description: Molecular characterization of a gene (LRG5) involved in blue light signal

A;Reference number: Z15399

A;Accession: T08179

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-640 <GLO>

A;Cross-references: EMBL:U73817; NID:g1644369; PID:g1644370

C;Genetics:

A;Gene: LRG5

Query Match 54.3%; Score 44; DB 2; Length 640;

Best Local Similarity 61.5%; Pred. No. 72;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 GSGAGTAAAGT 14

|||||:|||||

Db 440 GGGGVAVVAGT 452

RESULT 13

JC2465

copper-transporting ATPase (EC 3.6.1.-) HRA-2 - Enterobacteriaceae spp.

C;Species: Enterobacteriaceae spp.

C;Date: 16-Oct-1998 #sequence_revision 16-Oct-1998 #text_change 18-Jun-1999

C;Accession: JC2465

R;Frenor III., C.; Lin, W.; Andrews, N.C.

Biochem. Biophys. Res. Commun. 205, 1644-1650, 1994

A;Title: Novel bacterial P-type ATPases with histidine-rich heavy-metal-associated seque

A;Reference number: JC2464; MUID:95110304; PMID:7811248

A;Accession: JC2465

A;Molecule type: mRNA

A;Residues: 1-708 <GRE>

A;Cross-references: GB:U16659; NID:g643614; PIDN:AAA62114.1; PID:g643615

A;Experimental source: human small intestine cDNA library

A;Note: the source species is uncertain; the cloned sequence did not hybridize with huma

C;Function:

A;Description: copper transport coupled with ATP hydrolysis

C;Superfamily: Enterococcus copper-transporting ATPase copB; ATPase nucleotide-binding d

C;Keywords: copper transport; hydrolase; ion transport; phosphoprotein; transmembrane pr

F;83-99/Domain: transmembrane #status predicted <TM1>

F;114-130/Domain: transmembrane #status predicted <TM2>

F;115-455/Domain: ATPase transduction domain homology <ATT>

F;147-163/Domain: transmembrane #status predicted <TM3>

F;173-189/Domain: transmembrane #status predicted <TM4>

F;331-347/Domain: transmembrane #status predicted <TM5>

F;351-367/Domain: transmembrane #status predicted <TM6>

F;666-682/Domain: ATPase nucleotide-binding domain homology <ATN>

F;685-701/Domain: transmembrane #status predicted <TM7>

F;267/Active site: Glu #status predicted

F;411/Active site: Asp (aspartylphosphate intermediate) #status predicted

Query Match 54.3%; Score 44; DB 1; Length 708;

Best Local Similarity 62.5%; Pred. No. 79;

Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 GSGAGTAAAGTNNI 17

|||||:|||||

Db 618 GAGTQVAAADTVLI 633

RESULT 14

A36068

major ampullate fibroin protein - orb spider (Nephila clavipes) (fragment)

C;Species: Nephila clavipes

C;Date: 08-Mar-1991 #sequence_revision 13-Jan-1993 #text_change 09-Sep-1997

C;Accession: A36068

R;Xu, M.; Lewis, R.V.

Proc. Natl. Acad. Sci. U.S.A. 87, 7120-7124, 1990

A;Title: Structure of a protein superfiber: spider dragline silk.

A;Reference number: A36068; MUID:90384959; PMID:2402494

A;Accession: A36068

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-718 <XUA>

A;Cross-references: GB:M37137; NID:g159711; PID:g159712

A;Note: the authors translated the codon GGT for residue 292 as Gln, GTA for residue 374

Query Match 54.3%; Score 44; DB 2; Length 718;
 Best Local Similarity 71.4%; Pred. No. 80;
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GSGAGIAQAAAGTV 15
 | | | | |
 Db 557 GQGAGAAAAAAGGV 570

RESULT 15
 D82324
 conserved hypothetical protein VC0429 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
 C;Species: Vibrio cholerae
 C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C;Accession: D82324
 R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, J.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A;Reference number: A82035; MUID:20406833; PMID:10952301
 A;Accession: D82324
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-854 <HEI>
 A;Cross-references: GB:AE004130; GB:AE003852; NID:g9654845; PIDN:AAF93602.1; GSPDB:GN001
 A;Experimental source: serogroup O1; strain N16961; biotype El Tor
 C;Genetics:
 A;Gene: VC0429
 A;Map position: 1

Query Match 54.3%; Score 44; DB 2; Length 854;
 Best Local Similarity 75.0%; Pred. No. 95;
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GAGIAQAAAGTV 15
 | | | | |
 Db 515 GIGVATAAAGTV 526

Search completed: March 10, 2004, 12:12:26
 Job time : 10.1538 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:51:36 ; Search time 5.23077 Seconds
(without alignments)
169.228 Million cell updates/sec

Title: US-10-044-703-49

Perfect score: 81

Sequence: 1 TGSGAGIAQAAAGTVNI 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	81	100.0	374	1	PST1_MYCTU
2	46	56.8	364	1	PSTS_XYLFA
3	46	56.8	364	1	PSTS_XYLFT
4	44	54.3	300	1	YQGG_BACSU
5	44	54.3	416	1	R23B_MOUSE
6	44	54.3	747	1	SPDI_NSPCL
7	43	53.1	176	1	HSLV_PSEAE
8	43	53.1	598	1	VG9_EPR03
9	42	51.9	254	1	KDUD_BACSU
10	42	51.9	314	1	HCDH_HUMAN
11	42	51.9	314	1	HCDH_MOUSE
12	42	51.9	314	1	HCDH_PIG
13	42	51.9	314	1	HCDH_RAT
14	42	51.9	487	1	HOXH_ALCEU
15	42	51.9	700	1	DTNE_MOUSE
16	42	51.9	834	1	5HTA_DROME
17	42	51.9	1174	1	CIKE_DROME
18	42	51.9	3010	1	POLG_HCVBK
19	42	51.9	3010	1	POLG_HCVJA
20	42	51.9	3010	1	POLG_HCVTW
21	41	50.6	146	1	FLAV_DDSGI
22	41	50.6	278	1	Y4LA_RHSN
23	41	50.6	281	1	YL46_BRAJA
24	41	50.6	383	1	MANA_STRLI
25	41	50.6	916	1	HMDH_DROME
26	41	50.6	3010	1	POLG_HCVJT
27	40	49.4	152	1	Y4JB_RHSN
28	40	49.4	278	1	UPK1_STRCO
29	40	49.4	351	1	PSTS_RHILO
30	40	49.4	393	1	FLAA_PSEAE
31	40	49.4	438	1	YD98_RHIME
32	40	49.4	548	1	TOXA_COCCA
33	40	49.4	725	1	ATU_DROME

34	40	49.4	1047	1	ESG1_HUMAN
35	40	49.4	3011	1	POLG_HCVI
36	39.5	48.8	66	1	KINI_ARATH
37	39.5	48.8	637	1	SYT_NEIMA
38	39	48.1	225	1	MTDR_METKA
39	39	48.1	297	1	MTDB_METEX
40	39	48.1	342	1	HUPV_AZOCH
41	39	48.1	359	1	DHOM_YEAST
42	39	48.1	376	1	DHOM_SCHPO
43	39	48.1	388	1	SUCC_STAAM
44	39	48.1	388	1	SUCC_STAAP
45	39	48.1	443	1	COXX_HUMAN

ALIGNMENTS

RESULT 1

ID	PST1_MYCTU	STANDARD;	PRT;	374 AA.
AC	P15712; O05868;			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-APR-1990 (Rel. 14, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Phosphate-binding protein 1 precursor (PBP-1) (PstS-1) (Protein antigen B) (PAB) (Antigen Ag78).			
GN	PSTSI OR PHOS1 OR RV0934 OR MT0961 OR MTCY08D9.05C.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
[1]	SEQUENCE FROM N.A.			
RP	MEDLINE=9307568; PubMed=2545626;			
RA	Andersen A.B., Hansen E.B.;			
RT	"Structure and mapping of antigenic domains of protein antigen b, a 38,000-molecular-weight protein of Mycobacterium tuberculosis.";			
RL	Infect. Immun. 57:2481-2488(1989).			
[2]	SEQUENCE FROM N.A.			
RP	STRAIN=H37RV;			
RC	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";			
RL	Nature 393:537-544(1998).			
[3]	SEQUENCE FROM N.A.			
RP	STRAIN=CDC 1551 / Oshkosh;			
RC	MEDLINE=22206494; PubMed=12218036;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Emdinova M., Salzberg S.L., Bishai W., Uffordback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;			
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains.";			
RL	J. Bacteriol. 184:5479-5490(2002).			
CC	-!- FUNCTION: Required for binding-protein-mediated phosphate transport.			
CC	-!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor (Probable).			
CC	-!- SIMILARITY: Belongs to the pstS family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -			

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M30046; AAA25374.1; -
 CC EMBL; 295209; CAB08484.1; -
 CC EMBL; AE006981; AAK45208.1; -
 CC PIR; F70584; F70584.
 CC HSP; P06128; 11XH.
 CC TIGR; M0961; -
 CC TubercuList; RV0934; -
 CC InterPro; IPR000437; Prok lipoprot_S.
 CC InterPro; IPR006059; SBP_bac_1.
 CC Pfam; PF01547; SBP_bac_1; 1.
 CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 CC Phosphate transport; Transport; Membrane; Lipoprotein; Signal;
 KW Antigen; Complete proteome; Palmitate.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 374 PHOSPHATE-BINDING PROTEIN 1.
 FT LIPID 24 24 N-palmitoyl cysteine (Potential).
 FT LIPID 24 24 S-diacylglycerol cysteine (Potential).
 SQ SEQUENCE 374 AA; 38243 MW; 633496819F38AA CRC64;

Query Match 100.0%; Score 81; DB 1; Length 374;
 Best Local Similarity 100.0%; Pred. No. 0.00022;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGSAGAGIAQAAGTGVNI 17
 |||||
 DB 86 TGSAGAGIAQAAGTGVNI 102

RESULT 2

ID PSTS_XYLFA STANDARD; PRT; 364 AA.
 AC Q9P8K3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phosphate-binding periplasmic protein precursor (PBP).
 GN PPS OR XF2141.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 CC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9a5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvares R., Alves L.M.C., Araya J.B., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Facincini L.P., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincini A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krueger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

RA da Silva J.P., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 406:151-159(2000).
 CC -!- FUNCTION: Required for binding-protein-mediated phosphate
 CC transport (By similarity).
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -!- SIMILARITY: Belongs to the pps family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AE004028; AAF84940.1; -
 CC PIR; G82595; G82595.
 CC HSP; P06128; 11XH.
 CC InterPro; IPR000437; Prok lipoprot_S.
 CC InterPro; IPR006059; SBP_bac_1.
 CC Pfam; PF01547; SBP_bac_1; 1.
 CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 CC Phosphate transport; Transport; Periplasmic; Signal;
 KW Complete proteome.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 364 PHOSPHATE-BINDING PERIPLASMIC PROTEIN.
 SQ SEQUENCE 364 AA; 38736 MW; 8609CFAA159D4277 CRC64;

Query Match 56.8%; Score 45; DB 1; Length 364;

Best Local Similarity 60.0%; Pred. No. 17;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 GSGAGIAQAAGTGVN 16
 |||||
 DB 86 GSGAGIAQAAGTGVN 100

RESULT 3

PSTS_XYLFT
 ID PSTS_XYLFT STANDARD; PRT; 364 AA.
 AC Q87C91;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phosphate-binding periplasmic protein precursor (PBP).
 GN PPS OR PD1202.
 OS Xylella fastidiosa (strain Temecula / ATCC 700964).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 CC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=183190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22421331; PubMed=12533478;
 RA Van Sluys M.A., de Oliveira M.C., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
 RA Miyaki C.Y., Furian L.R., Machado M.A., Ferro M.I.T., da Silva F.R.,
 RA Takita M.A., Lemos E.G.M., Lemos M.V.F., El-Dorri H., Tsai S.M.,
 RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
 RA Carier H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
 RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.S.,
 RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
 RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
 RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
 RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
 RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
 RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
 RA Kitajima J.P.;
 RT "Comparative analyses of the complete genome sequences of Pierce's
 RT disease and citrus variegated chlorosis strains of Xylella
 RT fastidiosa.";

RL J. Bacteriol. 185:1018-1026(2003).

CC -!- FUNCTION: Required for binding-protein-mediated phosphate transport (by similarity).

CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).

CC -!- SIMILARITY: Belongs to the pstS family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR ENBL; AAC12557; AAC29053.1; --

DR InterPro: IPR000437; Prok_lipo_prot_S.

DR Pfam: PF01547; SBP_bac_1.1.

DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.

KW Phosphate transport; Transport; Periplasmic; Signal;

FT SIGNAL 1 18 POTENTIAL.

FT CHAIN 19 364 PHOSPHATE-BINDING PERIPLASMIC PROTEIN.

SQ SEQUENCE 364 AA; 38667 MW; 4499D9C3B10F4033 CRC64;

Query Match 56.8%; Score 46; DB 1; Length 364;

Best Local Similarity 60.0%; Pred. No. 17;

Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSGAGTAAQAAGTVN 16

DB 86 GSGGGIAQKAATID 100

RESULT 4

YQGG_BACSU STANDARD; PRT; 300 AA.

AC F46338; --

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Probable ABC transporter binding protein yqgg precursor.

GN YQGG OR BSU24990.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168 / JH642;

RX MEDLINE=96349106; PubMed=8760913;

RA Takemaru K.-I., Mizuno M., Kobayashi Y.;

RT "A Bacillus subtilis gene cluster similar to the Escherichia coli phosphate-specific transport (pst) operon: evidence for a tandemly arranged pstB gene.";

RL Microbiology 142:2017-2020(1996).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168 / JH642;

RX MEDLINE=97124195; PubMed=8969508;

RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M., Kobayashi Y.;

RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of the Bacillus subtilis genome containing the skin element and many sporulation genes.";

RL Microbiology 142:3103-3111(1996).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Arazevedo V., Bertsier M.G., Bessieres P., Bolotin A., Borchert S., Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,

Choi S.K., Codani J.J., Comerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabret C., Fertari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Guiseppe G., Guy B.J., Hega K., Haeck J., Harwood C.R., Henaut A., Hilbert H., Holbappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kaahara Y., Klaerr-Bianchard M., Klein C., Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moser D., Nakai S., Noback M., Neone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivalta C., Rocha E., Roche B., Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;

RT "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.";

RL Nature 390:249-256(1997).

CC -!- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM YQGGHIJK.

CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor (Probable).

CC -!- SIMILARITY: Belongs to the pstS family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR ENBL; D58414; BAA09581.1; --

DR ENBL; D84432; BAA12510.1; --

DR ENBL; Z99116; CAB14429.1; --

DR PIR; A69956; A69956.

DR Subtilist; BG11375; YQGG.

DR InterPro: IPR000437; Prok_lipo_prot_S.

DR Pfam: PF01547; SBP_bac_1.1.

DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.

KW Hypothetical protein; Transport; Lipoprotein; Membrane; Signal;

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 300 PROBABLE ABC TRANSPORTER BINDING PROTEIN

FT YQGG.

FT LIPID 22 22 N-palmitoyl cysteine (Probable).

FT LIPID 22 22 S-diacylglycerol cysteine (Probable).

SQ SEQUENCE 300 AA; 31684 MW; 9DBA6090947A277B CRC64;

Query Match 54.3%; Score 44; DB 1; Length 300;

Best Local Similarity 50.0%; Pred. No. 28;

Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 GSGAGTAAQAAGTVNI 17

DB 80 GSGTGLSQVSEGAQVI 95

RESULT 5

R23B_MOUSE

ID R23B_MOUSE STANDARD; PRT; 416 AA.

AC P54728;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1986 (Rel. 34, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DT UV excision repair protein RAD23 homolog B (MHR23B) (XP-C repair
DE complementing complex 58 kDa protein) (P58).
GN RAD23B OR MHR23B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Testis;
RX MEDLINE=96403997; PubMed=9808275;
RA van der Spek P.J., Visser C.E., Hanoka F., Smit B.,
RA Hagemeijer A., Boutsma D., Hoeijmakers J.H.J.;
RT "Cloning, comparative mapping, and RNA expression of the mouse
RT homologues of the Saccharomyces cerevisiae nucleotide excision repair
RT gene RAD23.";
RL Genomics 31:20-27(1996).
CC -!- FUNCTION: Involved in DNA excision repair. May play a part in DNA
CC damage recognition and/or in altering chromatin structure to allow
CC access by damage-processing enzymes.
CC -!- SUBUNIT: HETERODIMER OF A 125 kDa SUBUNIT (P125) AND OF A
CC 58 kDa SUBUNIT (P58).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Contains 1 ubiquitin-like domain.
CC -!- SIMILARITY: Contains 2 UBA domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; X92411; CRA63146.1; -;
CC HSBP; P54725; LDV0.

DR MGD; MGI:105128; Rad23b.
DR InterPro; IPR004806; Rad23.
DR InterPro; IPR006636; SRI1.
DR InterPro; IPR000449; UBA domain.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00240; Ubiquitin; 1.
DR SMART; SM00727; STI1; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00213; UBQ; 1.
DR TIGRFAMs; TIGR00601; rad23; 1.
DR PROSITE; PS50030; UBA; 2.
DR PROSITE; PS50053; UBIQUITIN 2; 1.
KW DNA damage; DNA repair; Nuclear protein; Repeat.
FT DOMAIN 1 79 UBIQUITIN-LIKE.
FT DOMAIN 188 228 UBA 1.
FT DOMAIN 371 411 UBA 2.
FT DOMAIN 255 261 POLY-ALA.
FT DOMAIN 262 270 POLY-THR.
FT DOMAIN 336 355 POLY-GLY.
SQ SEQUENCE 416 AA; 43516 MW; 13E0245ACD982205 CRC64;

Query Match 54.3%; Score 44; DB 1; Length 416;
Best Local Similarity 53.3%; Pred. No. 37;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 GSGAGIAAAGTGN 16
Db 351 GGGGIAAGSGHWN 365

RESULT 6

SPD1_NEPCL STANDARD; PRT; 747 AA.
ID _SPD1_NEPCL
AC P19837;

DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Spidroin 1 (Dragline silk fibroin 1) (Pragmat).
DE Nephila clavipes (Orb spider).
OS Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90384959; PubMed=2402494;
RA Xu M., Lewis R.V.;
RT "Structure of a protein superfiber: spider dragline silk.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124 (1990).
RN [2]
RP SEQUENCE OF 653-747 FROM N.A.
RX MEDLINE=94165058; PubMed=8120021;
RA Beckwith R., Arcidiacono S.;
RT "Sequence conservation in the C-terminal region of spider silk
RT proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and
RT Araneus bicentenarius (Araneidae).";
RL J. Biol. Chem. 269:8661-8663(1994).
CC -!- FUNCTION: Spiders major ampullate silk possesses unique
CC pseudocrystalline regions of antiparallel beta-sheet interspersed
CC with elastic amorphous segments.
CC -!- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- DOMAIN: Highly repetitive protein characterized by regions of
CC polyalanine and glycine-rich repeating units.
CC -!- SIMILARITY: Belongs to the silk fibroin family.
CC -!- DATABASE: NAME=Protein Spotlight;
CC NOTE=Issue 24 of July 2002;
CC WWW=<http://www.expasy.org/spotlight/articles/spl024.html>.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; M37137; AAA29380.1; -;
CC EMBL; U03848; AAB60212.1; -;
CC PIR; A36068; A36068.
CC Silk; Repeat.

FT NON_TER 1 655 25 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 1 25
FT REPEAT 1 25
FT REPEAT 26 38
FT REPEAT 39 66
FT REPEAT 67 96
FT REPEAT 97 130
FT REPEAT 131 158
FT REPEAT 159 191
FT REPEAT 192 204
FT REPEAT 205 235
FT REPEAT 236 262
FT REPEAT 263 292
FT REPEAT 293 305
FT REPEAT 306 333
FT REPEAT 334 360
FT REPEAT 361 394
FT REPEAT 395 424
FT REPEAT 425 458
FT REPEAT 459 485
FT REPEAT 486 512
FT REPEAT 513 525
FT REPEAT 526 555
FT REPEAT 556 582
FT REPEAT 583 612
FT REPEAT 613 642

```

FT REPEAT      643 655 25.
FT CONFLICT    662 662 V -> L (IN REF. 1).
FT CONFLICT    662 662 S -> T (IN REF. 1).
FT CONFLICT    672 672 NPGLSGCDVLIQALLEVLSALIQILGSSSTGVNYSAGQA
FT CONFLICT    695 747 TQIVGQSVYQAL -> ILVFLDWNSSPKFSRLFLILRS
FT (IN REF. 1).
SQ SEQUENCE 747 AA; 60528 MW; 850E44B0D649E012 CRC64;

Query Match      54.3%; Score 44; DB 1; Length 747;
Best Local Similarity 71.4%; Pred. No. 61;
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GSGAGIAQAAGTV 15
Db 557 GQAGAAAAAGGV 570

RESULT 7
HSLV_PSBAE
ID HSLV_PSBAE STANDARD; PRT; 176 AA.
AC Q9HUC6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ATP-dependent protease hslv (EC 3.4.25.-).
GN HSLV OR PA5053;
OS Pseudomonas aeruginosa;
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey M.J., Brinkman F.S.L., Hurnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -!- FUNCTION: Protease subunit of a proteasome-like degradation
CC complex (By similarity).
CC -!- SUBUNIT: A double ring-shaped homohexamer of hslv is capped on
CC each side by a ring-shaped hslu homohexamer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family T1B. Hslv subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AEO04918; AAC08438.1; -
CC DR PIR; D83015; D83015.
CC DR HSPP; P31059; 1E94.
CC DR MEROPS; T01.006; -.
CC DR HAMAP; MF_00248; -.
CC DR InterPro; IPR001353; Peptidase_T1.
CC DR Pfam; PF00227; proteasome; 1.
CC DR Hydrolase; Protease; Threonine protease; Complete proteome.
CC INIT MET 0
CC ACT SITE 1 BY SIMILARITY.
CC ACT SITE 1 BY SIMILARITY.
CC SEQUENCE 176 AA; 18654 MW; 291076FP451E5B09 CRC64;

Query Match      53.1%; Score 43; DB 1; Length 176;
Best Local Similarity 81.8%; Pred. No. 24;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 2 GSGAGIAQAAA 12
Db 123 GSGGGAFAQAAA 133

RESULT 8
VG9_BPB03
ID VG9_BPB03 STANDARD; PRT; 598 AA.
AC Q37E90;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Tail protein (late protein GP9).
GN GP9.
OS Bacteriophage B103.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC Phi-29-like viruses.
OX NCBI_TaxID=10778;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98019084; PubMed=9358052;
RA Pecenkova T., Benes V., Paces J., Vlcek C., Paces V.;
RT "Bacteriophage B103: complete DNA sequence of its genome and
RT relationship to other Bacillus phages."
RL Gene 199:157-163(1997).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X99260; CAA67657.1; -
CC DR LATE protein.
CC SQ SEQUENCE 598 AA; 67437 MW; A13DBB185027E5F0 CRC64;

Query Match      53.1%; Score 43; DB 1; Length 598;
Best Local Similarity 50.0%; Pred. No. 69;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TGSGAGIAQAAAAGTVN 16
Db 451 TGSAVGAVASSATGMVS 466

RESULT 9
KDUD_BACSU
ID KDUD_BACSU STANDARD; PRT; 254 AA.
AC P50842;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 2-deoxy-D-glucuronate 3-dehydrogenase (EC 1.1.1.125) (2-keto-3-
DE deoxyglucuronate oxidoreductase).
GN KDUD OR BSU22140.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / Marburg;
RX MEDLINE=96349105; PubMed=8760912;
RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,
RA Serron P.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RT the serA and kgd loci cloned in a yeast artificial chromosome."
RL Microbiology 142:2005-2016(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;

```

```

RX MEDLINE=9804033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertsch M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut L.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maui C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Ncone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Piescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scallan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256 (1997).
CC -!- CATALYTIC ACTIVITY: 2-deoxy-D-gluconate + NAD(+) = 3-dehydro-2-
CC deoxy-D-gluconate + NADH.
CC -!- PATHWAY: Pectin degradation.
CC -!- SIMILARITY: Belongs to the short-chain dehydrogenases/reductases
CC (SDR) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L47838; AAB38476.1; .
CC EMBL; Z99115; CAB14131.1; .
CC PIR; D69648; D69648.
CC HSSP; 070351; 136W.
CC Subtilist; EG1400; kdud.
CC InterPro; IPR002198; ADH_short.
CC Pfam; PF00106; adh_short; 1.
CC PRINTS; PRO0080; SDRFAMILY.
CC PROSITE; PS00061; ADH_SHORT; 1.
KW Oxidoreductase; NAD; Complete proteome.
FT NP_BIND 16 40 NAD (BY SIMILARITY).
FT ACT_SITE 159 159 BY SIMILARITY.
SQ SEQUENCE 254 AA; 27185 MW; 245C57CF5CA46597 CRC64;

Query Match 51.9%; Score 42; DB 1; Length 254;
Best Local Similarity 69.2%; Pred. No. 46;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGSGAGIAQAAG 13
DB 22 TGIGGQIAKALAG 34

RESULT 10
HCDH_HUMAN
ID HCDH_HUMAN STANDARD; PRT; 314 AA.
AC Q16836; O00753;

```

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Short chain 3-hydroxyacyl-CoA dehydrogenase, mitochondrial precursor
DE (EC 1.1.1.35) (HCDH) (Medium and short chain L-3-hydroxyacyl-coenzyme
DE A dehydrogenase).
GN HADHSC OR SCHAD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=96268746; PubMed=8687483;
RA Vredendaal P.J.C.M., van den Berg I.E.T., Malingre H.E.M.,
RA Strobants A.K., Oldeweghuis D.E.M., Berger R.;
RT "Human short-chain L-3-hydroxyacyl-CoA dehydrogenase: cloning and
RT characterization of the coding sequence."
RL Biochem. Biophys. Res. Commun. 223:718-723 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX O'Brien L.K., Sims H.F., Strauss A.W.;
RT "Human short chain L-3-hydroxyacyl-CoA dehydrogenase."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=22368257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Roshylyki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [4]
RP SEQUENCE OF 7-314 FROM N.A.
RC TISSUE=Skeletal muscle;
RA Shi Y., Samuel S.J., Lee W., Yu C.H., Zhang W., Lachaal M., Jung C.Y.;
RT "Cloning of a L-3-hydroxyacyl CoA dehydrogenase that binds to GLUT4
RT glucose transporter cytoplasmic C-terminus: possible crosstalk between
RT glucose transport and fatty acid metabolism."
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 13-314.
RX MEDLINE=99249789; PubMed=10231530;
RA Barycki J.J., O'Brien L.K., Bratt J.M., Zhang R., Sanishvili R.,
RA Strauss A.W., Banaszak L.J.;
RT "Biochemical characterization and crystal structure determination of
RT human heart short chain L-3-hydroxyacyl-CoA dehydrogenase provide
RT insights into catalytic mechanism."
RL Biochemistry 38:5786-5798 (1999).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 13-314.
RX MEDLINE=20418095; PubMed=10840044;
RA Barycki J.J., O'Brien L.K., Strauss A.W., Banaszak L.J.;
RT "Sequestration of the active site by interdomain shifting.
RT Crystallographic and spectroscopic evidence for distinct
RT conformations of L-3-hydroxyacyl-CoA dehydrogenase."

```


J. Biol. Chem. 275:27196-27196(2000).

-1- FUNCTION: Plays an essential role in the mitochondrial beta-oxidation of short chain fatty acids. Exerts it highest activity toward 3-hydroxybutyryl-CoA.

-1- CATALYTIC ACTIVITY: (S)-3-hydroxyacyl-CoA + NAD(+) = 3-oxoacyl-CoA + NADH.

-1- PATHWAY: Fatty acid beta-oxidation cycle; step 3.

-1- SUBUNIT: Homodimer.

-1- TISSUE SPECIFICITY: Mitochondrial matrix.

-1- TISSUE SPECIFICITY: Expressed in liver, kidney, pancreas, heart and skeletal muscle.

-1- SIMILARITY: Belongs to the 3-hydroxyacyl-CoA dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X96752; CAA65528.1; -

EMBL; AF095703; AAD13581.1; -

EMBL; BC000306; AAH00306.1; -

EMBL; AF001902; AAB54008.1; -

PIR; JC4879; JC4879.

PDB; 1FOY; 01-SEP-00.

PDB; 1F14; 27-SEP-00.

PDB; 3HAD; 12-JAN-00.

PDB; 1ILO; 07-NOV-01.

Genew; HGNC:4804; HADHSC.

GK; Q16836; -

MM; 601609; -

GO; GO:0005739; C:mitochondrion; TAS.

GO; GO:0003857; F:3-hydroxyacyl-CoA dehydrogenase activity; TAS.

InterPro; IPR006180; 3HCDH.

InterPro; IPR006108; 3HCDH C.

InterPro; IPR006176; 3HCDH N.

InterPro; IPR008927; 6DGDH C-like.

InterPro; IPR00205; NAD_BS.

Pfam; PF00725; 3HCDH; 1.

Pfam; PF02737; 3HCDH N; 1.

PROSITE; PS00067; 3HCDH; 1.

Fatty acid metabolism; Oxidoreductase; NAD; Mitochondrion;

Transit peptide; 3D-structure.

TRANSIT 1 12

CHAIN 13 314

MITOCHONDRION

SHORT CHAIN 3-HYDROXYACYL-CoA

DEHYDROGENASE.

GENERAL BASE.

Q -> H (IN REF. 1).

ACT SITE 170 170

CONFLICT 152 152

STRAND 29 33

HELIX 37 48

TURN 49 50

STRAND 52 56

HELIX 60 79

TURN 80 81

HELIX 86 98

TURN 99 99

STRAND 100 103

HELIX 106 109

TURN 110 111

STRAND 115 118

HELIX 124 134

TURN 135 137

TURN 140 141

STRAND 143 146

HELIX 153 157

TURN 158 159

HELIX 163 165

STRAND 166 171

TURN 175 177

STRAND 180 184

FT TURN 187 188

FT HELIX 191 203

FT TURN 204 205

FT STRAND 207 211

FT TURN 215 218

FT HELIX 219 235

FT TURN 236 237

FT HELIX 241 252

FT HELIX 258 265

FT HELIX 267 279

FT TURN 280 283

FT HELIX 285 287

FT TURN 291 298

FT TURN 299 300

FT STRAND 303 303

FT TURN 304 307

FT STRAND 308 308

FT STRAND 312 312

SQ SEQUENCE 314 AA; 34277 MW; 825795F8F972CFF16 CRC64;

Query Match 51.9%; Score 42; DB 1; Length 314;

Best Local Similarity 81.8%; Pred. No. 55;

Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GAGTAQAAGT 14

DB 39 GAGTAQAAGT 49

RESULT 11

HCDH MOUSE

ID HCDH MOUSE STANDARD; PRT; 314 AA.

AC Q61425; Q925U9;

DT 01-NOV-1997 (Rel. 35, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Short chain 3-hydroxyacyl-CoA dehydrogenase, mitochondrial precursor (EC 1.1.1.35) (HCDH) (Medium and short chain L-3-hydroxyacyl-coenzyme A dehydrogenase).

GN HADHSC OR SCHAD OR HADH OR MSCHAD.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1] _taxid=10090;

RP SEQUENCE FROM N.A.

RX MEDLINE=95369712; PubMed=7642117;

RA Nomura M., Takihara Y., Shimada K.;

RT "Isolation of a cDNA clone encoding mouse 3-hydroxyacyl CoA dehydrogenase."

RL Gene 160:309-310(1995).

RN [2]

RP SEQUENCE FROM N.A.

RA O'Brien L.K., Sims H.F., Strauss A.W.;

RT "Mouse medium and short chain L-3-hydroxyacyl-coenzyme A dehydrogenase gene."

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Plays an essential role in the mitochondrial beta-oxidation of short chain fatty acids. Exerts it highest activity toward 3-hydroxybutyryl-CoA.

CC -1- CATALYTIC ACTIVITY: (S)-3-hydroxyacyl-CoA + NAD(+) = 3-oxoacyl-CoA + NADH.

CC -1- PATHWAY: Fatty acid beta-oxidation cycle; step 3.

CC -1- SUBUNIT: Homodimer (By similarity).

CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.

CC -1- SIMILARITY: Belongs to the 3-hydroxyacyl-CoA dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

or send an email to license@isb-sib.ch).

EMBL; D29639; BAA06122.1; ALT INIT.
 EMBL; AF375597; AAKS4642.1; --
 EMBL; AF375596; AAKS4642.1; JOINED.
 PIR; JC4210; JC4210.
 HSP; P00348; 3HDH.
 SWISS-2DPAGE; Q61425; MOUSE.
 MGD; MGI:96009; Madhes.
 InterPro; IPR006180; 3HCDH.
 InterPro; IPR006108; 3HCDH.C.
 InterPro; IPR006176; 3HCDH.N.
 InterPro; IPR008927; 6GDH_C-like.
 InterPro; IPR000205; NAD_BS.
 Pfam; PF00725; 3HCDH; 1.
 Pfam; PF02737; 3HCDH.N; 1.
 PROSITE; PS00067; 3HCDH; FALSE NEG.
 Fatty acid metabolism; Oxidoreductase; NAD; Mitochondrion;
 Transit peptide.
 TRANSIT 1 12 MITOCHONDRION (BY SIMILARITY).
 CHAIN 13 314 SHORT CHAIN 3-HYDROXYACYL-COA
 DEHYDROGENASE.
 ACT_SITE 170 170 GENERAL BASE (BY SIMILARITY).
 CONFLICT 111 111 H -> D (IN REF. 1).
 CONFLICT 211 211 C -> S (IN REF. 1).
 SEQUENCE 314 AA; 34464 MW; 366A4C075F708BC1 CRC64;

Query Match 51.9%; Score 42; DB 1; Length 314;
 Best Local Similarity 81.8%; Pred. No. 55;
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GAGIAQAAGT 14
 Db 39 GAGIAQVAAT 49

RESULT 12
 HCDH.PIG
 ID HCDH.PIG STANDARD; PRT; 314 AA.
 AC P00348; Q9XS66;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Short chain 3-hydroxyacyl-CoA dehydrogenase, mitochondrial precursor
 DE (EC 1.1.1.35) (HCDH) (Medium and short chain L-3-hydroxyacyl-coenzyme
 DE A dehydrogenase).
 GN HADHSC OR SCHAD OR HAD.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=98256028; PubMed=9593854;
 RA He X.-Y., Yang S.-Y.;
 RT "Molecular cloning, expression in Escherichia coli, and
 RT characterization of a novel L-3-hydroxyacyl coenzyme A dehydrogenase
 RT from pig liver.";
 RL Biochim. Biophys. Acta 1392:119-126(1998).
 RN [2]
 RP SEQUENCE OF 13-314.
 RC TISSUE=Heart;
 RX MEDLINE=81004379; PubMed=7409145;
 RA Bitar K.G., Perez-Aranda A., Bradshaw R.A.;
 RT "Amino acid sequence of L-3-hydroxyacyl CoA dehydrogenase from pig
 RT heart muscle.";
 RL FEBS Lett. 116:196-198(1980).
 RN [3]
 RP REVISIONS TO 16 AND 21.
 RA Pang J.-K., Bradshaw R.A.;
 RL Submitted (OCT-1982) to the PIR data bank.
 RN [4]

X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 MEDLINE=8068574; PubMed=3479790;
 Birkhoff J.J., Holden H.M., Hamlin R., Xuong N.H., Banaszak L.J.;
 "Structure of L-3-hydroxyacyl-coenzyme A dehydrogenase: Preliminary
 chain tracing at 2.8-A resolution";
 Proc. Natl. Acad. Sci. U.S.A. 84:8262-8266(1987).
 [5]
 X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 TISSUE=Heart;
 MEDLINE=20014023; PubMed=10548046;
 Barycki J.J., O'Brien L.K., Birkhoff J.J., Strauss A.W.,
 Banaszak L.J.;
 "Pig heart short chain L-3-hydroxyacyl-CoA dehydrogenase revisited:
 sequence analysis and crystal structure determination.";
 Protein Sci. 8:2010-2018(1999).
 [6]
 SHOWS THAT HEART AND LIVER ENZYMES ARE IDENTICAL.
 MEDLINE=99165583; PubMed=10064895;
 He X.-Y., Zhang G., Blecha P., Yang S.-Y.;
 "Identity of heart and liver L-3-hydroxyacyl coenzyme A
 dehydrogenase.";
 Biochim. Biophys. Acta 1437:119-123(1999).
 -!- FUNCTION: Plays an essential role in the mitochondrial beta-
 oxidation of short chain fatty acids. Exerts it highest activity
 toward 3-hydroxybutyryl-CoA.
 -!- CATALYTIC ACTIVITY: (S)-3-hydroxyacyl-CoA + NAD(+) = 3-oxoacyl-CoA
 + NADH
 -!- PATHWAY: Fatty acid beta-oxidation cycle; step 3.
 -!- SUBUNIT: Homodimer.
 -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
 -!- SIMILARITY: Belongs to the 3-hydroxyacyl-CoA dehydrogenase family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; AF027652; RAD20939.1; --
 PIR; T48866; T48866.
 PDB; 3HDH; 08-OCT-99.
 InterPro; IPR006180; 3HCDH.
 InterPro; IPR006108; 3HCDH.C.
 InterPro; IPR006176; 3HCDH.N.
 InterPro; IPR008927; 6GDH_C-like.
 InterPro; IPR000205; NAD_BS.
 Pfam; PF00725; 3HCDH; 1.
 Pfam; PF02737; 3HCDH.N; 1.
 PROSITE; PS00067; 3HCDH; 1.
 Fatty acid metabolism; Oxidoreductase; NAD; Mitochondrion;
 Transit peptide; 3D-structure.
 TRANSIT 1 12 MITOCHONDRION (BY SIMILARITY).
 CHAIN 13 314 SHORT CHAIN 3-HYDROXYACYL-COA
 DEHYDROGENASE.
 ACT_SITE 170 170 GENERAL BASE (BY SIMILARITY).
 CONFLICT 108 108 S -> A (IN REF. 2).
 CONFLICT 122 122 E -> EQLKVVGE (IN REF. 2).
 CONFLICT 172 174 FNP -> N (IN REF. 2).
 STRAND 29 33
 HELIX 37 48
 TURN 49 50
 STRAND 52 56
 HELIX 62 81
 TURN 82 84
 HELIX 86 98
 TURN 99 99
 STRAND 100 103
 HELIX 106 108
 TURN 109 109
 STRAND 115 118
 HELIX 124 137

FT STRAND 143 146
 FT HELIX 153 159
 FT HELIX 163 165
 FT STRAND 166 170
 FT TURN 175 177
 FT STRAND 180 185
 FT TURN 187 188
 FT STRAND 191 203
 FT HELIX 204 205
 FT TURN 207 212
 FT STRAND 215 218
 FT TURN 219 233
 FT STRAND 234 234
 FT HELIX 243 251
 FT HELIX 258 265
 FT TURN 267 271
 FT HELIX 272 277
 FT TURN 278 278
 FT TURN 291 295
 FT TURN 297 300
 FT TURN 304 306
 SQ SEQUENCE 314 AA; 34161 MW; 596CBFD227214C3B CRC64;

Query Match 51.9%; Score 42; DB 1; Length 314;
 Best Local Similarity 81.8%; Pred. No. 55;
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GAGIAQAAAGT 14
 |||||
 DB 39 GAGIAQVAAAT 49

RESULT 13
 HCDH RAT
 ID HCDH RAT STANDARD; PRT; 314 AA.
 AC Q9WVK7;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Short chain 3-hydroxyacyl-CoA dehydrogenase, mitochondrial precursor
 DE (EC 1.1.1.35) (HCDH) (Medium and short chain L-3-hydroxyacyl-coenzyme
 A dehydrogenase).
 GN HADHSC OR SCHAD OR HAD.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Heart muscle, and Liver;
 RX MEDLINE=99165583; PubMed=10064895;
 RA He X.-Y., Zhang G., Blecha F., Yang S.-Y.;
 RT "Identity of heart and liver L-3-hydroxyacyl coenzyme A
 dehydrogenase.";
 RL Biochim. Biophys. Acta 1437:119-123(1999).
 CC -!- FUNCTION: Plays an essential role in the mitochondrial beta-
 oxidation of short chain fatty acids. Exerts it highest activity
 toward 3-hydroxybutyryl-CoA.
 CC -!- CATALYTIC ACTIVITY: (S)-3-hydroxyacyl-CoA + NAD(+) = 3-oxoacyl-CoA
 + NADH.
 CC -!- PATHWAY: Fatty acid beta-oxidation cycle; step 3.
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -!- SIMILARITY: Belongs to the 3-hydroxyacyl-CoA dehydrogenase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; AF095449; RAD42162.1; -.
 DR HSSP; P00348; 3HDX.
 DR InterPro; IPR006180; 3HCDH.
 DR InterPro; IPR006108; 3HCDH_C.
 DR InterPro; IPR006176; 3HCDH_N.
 DR InterPro; IPR008927; 6DGDH_C-like.
 DR InterPro; IPR000205; NAD BS.
 DR Pfam; PF00725; 3HCDH; 1.
 DR Pfam; PF02737; 3HCDH_N; 1.
 DR PROSITE; PS00067; 3HCDH; FALSE NEG.
 KW Fatty acid metabolism; Oxidoreductase; NAD; Mitochondrion;
 KW Transit peptide.
 FT TRANSIT 1 12 MITOCHONDRION (BY SIMILARITY).
 FT CHAIN 13 314 SHORT CHAIN 3-HYDROXYACYL-COA
 DEHYDROGENASE.
 FT ACT SITE 170 170 GENERAL BASE (BY SIMILARITY).
 SQ SEQUENCE 314 AA; 34448 MW; B36518766D6C3828 CRC64;

Query Match 51.9%; Score 42; DB 1; Length 314;
 Best Local Similarity 81.8%; Pred. No. 55;
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GAGIAQAAAGT 14
 |||||
 DB 39 GAGIAQVAAAT 49

RESULT 14
 HOXH ALCEU
 ID HOXH ALCEU STANDARD; PRT; 487 AA.
 AC P22320;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE NAD-reducing hydrogenase hoxS beta subunit (EC 1.1.2.1.2).
 GN HOXH.
 OS Alcaligenes eutrophus (Ralstonia eutropha).
 OC Plasmid megaplasmid pHG1.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 OX NCBI_TaxID=510;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H16 / DSM 428 / ATCC 17699;
 RX MEDLINE=90264277; PubMed=2188945;
 RA Tran-Betcke A., Warnecke U., Bocker C., Zaborosch C., Friedrich B.;
 RT "Cloning and nucleotide sequences of the genes for the subunits of
 NAD-reducing hydrogenase of *Alcaligenes eutrophus* H16.";
 RL J. Bacteriol. 172:2920-2929(1990).
 RN [2]
 RP SEQUENCE OF 1-38.
 RC STRAIN=H16 / DSM 428 / ATCC 17699;
 RX MEDLINE=89231684; PubMed=2496982;
 RA Zaborosch C., Schneider K., Schlegel H.G., Kratzin H.;
 RT "Comparison of the NH₂-terminal amino acid sequences of the four non-
 identical subunits of the NAD-linked hydrogenases from *Nocardia opaca*
 1b and *Alcaligenes eutrophus* H16.";
 RL Eur. J. Biochem. 181:175-180(1989).
 CC -!- CATALYTIC ACTIVITY: H(2) + NAD(+) = H(+) + NADH.
 CC -!- COFACTOR: FMN and nickel.
 CC -!- SUBUNIT: Tetramer of an alpha and a gamma subunits (flavin-
 containing dimer), and a delta and a nickel-containing beta
 subunits (hydrogenase dimer).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the [NiFe]/[NiFeS] hydrogenase large
 subunit family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M55230; AAC06143.1; -.
CC PIR; D35385; D35385
CC InterPro; IPR001501; N4 hdl.
CC Pfam; PF00374; Nifese_Hases; 1; 1.
CC PROSITE; PS00507; N1_HGNASE_L1; 1.
CC PROSITE; PS00508; N1_HGNASE_L2; 1.
CC Oxidoreductase; NAD; Metal-binding; Nickel; Plasmid.
KW INIT MET 0
FT METAL 61 NICKEL (POTENTIAL).
FT METAL 64 NICKEL (POTENTIAL).
FT METAL 457 NICKEL (POTENTIAL).
FT METAL 460 NICKEL (POTENTIAL).
FT CONFLICT 29 V -> R (IN REF. 2).
FT CONFLICT 31 A -> T (IN REF. 2).
FT CONFLICT 36 V -> R (IN REF. 2).
SQ SEQUENCE 487 AA; 54754 MW; 639A4F6C9C05D3C4 CRC64;

Query Match 51.9%; Score 42; DB 1; Length 487;
Best Local Similarity 46.7%; Pred. No. 80;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGSGAGIAQAAAGTV 15
||| : : |||
DB 379 TGGVGVEPRGTL 393

RESULT 15
DTNE MOUSE STANDARD; PRT; 700 AA.
AC Q70555; O70563; Q9CTZ1.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Dystrobrevin beta (Beta-dystrobrevin) (DTN-B) (MDTN-B).
GN DTNB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CO NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=98200066; PubMed=9540997;
RA Puca A.A., Piluso V.N.G., Belisio A., Sampaolo S., Quaderi N.,
RA Rossi E., Di Iorio G., Ballabio A., Franco B.;
RT Identification and characterization of a novel member of the
RT dystrobrevin gene family."
RL FEBS Lett. 425:7-13(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX MEDLINE=98081858; PubMed=9419360;
RA Blake D.J., Nawrotzki R., Loh N.Y., Gorecki D.C., Davies K.E.;
RT "Beta-dystrobrevin, a member of the dystrophin-related protein
RT family."
RL Proc. Natl. Acad. Sci. U.S.A. 95:241-246(1998).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX MEDLINE=99018217; PubMed=9799833;
RA Loh N.Y., Ambrose H.J., Guay-Woodford L.M., Dasgupta S.,
RA Nawrotzki R.A., Blake D.J., Davies K.E.;
RT "Genomic organization and refined mapping of the mouse beta-
RT dystrobrevin gene."
RL Mamm. Genome 9:857-862(1998).
RN [4]
RP SEQUENCE OF 590-608 FROM N.A. (ISOFORM 2).
RC STRAIN=C57BL/6J; Tissue=Stomach;
RX MEDLINE=21085660; PubMed=11217851;
RA Arakawa T., Hara A., Fukunishi Y., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

```

```

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli P., Suzuki R., Tonita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baidarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bulc C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
RN [5]
RX INTERACTION WITH DYSTROBREVIN BINDING PROTEIN 1.
MEDLINE=21316514; PubMed=11316798;
RA Benson M.A., Newey S.E., Martin-Rendon E., Hawkes R., Blake D.J.;
RT "Dysbindin, a novel coiled-coil-containing protein that interacts with
RT the dystrobrevins in muscle and brain."
RL J. Biol. Chem. 276:24232-24241(2001).
CC -!- SUBUNIT: Interacts with dystrophin short form DP71 and syntrophins
CC -!- SUBUNIT: Interacts with syntrophin. Binds dystrobrevin binding
CC protein 1.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=O70585-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O70585-2; Sequence=VSP 004227, VSP 004228, VSP 004229;
-!- TISSUE SPECIFICITY: Expressed mainly in brain, kidney, liver and
CC lung. In brain expressed in neurons of the cortex and hippocampus.
CC -!- DOMAIN: The coiled-coil domain may mediate the interaction with
CC dystrophin.
CC -!- SIMILARITY: BELONGS TO THE DYSTROPHIN FAMILY. DYSTROBREVIN
CC SUBFAMILY.
CC -!- SIMILARITY: Contains 1 ZZ-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y15742; CAA75752.1; -.
CC EMBL; AJ003007; CAA05796.1; -.
CC EMBL; AJ010204; CAA09038.1; JOINED.
CC EMBL; AJ010205; CAA09038.1; JOINED.
CC EMBL; AJ010206; CAA09038.1; JOINED.
CC EMBL; AJ010207; CAA09038.1; JOINED.
CC EMBL; AJ010208; CAA09038.1; JOINED.
CC EMBL; AJ010209; CAA09038.1; JOINED.
CC EMBL; AJ010210; CAA09038.1; JOINED.
CC EMBL; AJ010211; CAA09038.1; JOINED.
CC EMBL; AJ010212; CAA09038.1; JOINED.
CC EMBL; AJ010213; CAA09038.1; JOINED.
CC EMBL; AJ010214; CAA09038.1; JOINED.
CC EMBL; AJ010215; CAA09038.1; JOINED.
CC EMBL; AJ010216; CAA09038.1; JOINED.
CC EMBL; AJ010217; CAA09038.1; JOINED.
CC EMBL; AJ010218; CAA09038.1; JOINED.
CC EMBL; AJ010219; CAA09038.1; JOINED.
CC EMBL; AJ010220; CAA09038.1; JOINED.
CC EMBL; AJ010221; CAA09038.1; JOINED.
CC EMBL; AK019068; -; NOT ANNOTATED_CDS.
CC MGD; MGI:1203728; Dnrb.
CC GO; GO:0045202; C:synaptic junction; IDA.
CC InterPro; IPR000433; Znf_ZZ.

```

```

DR Pfam: PF00569; ZZ; 1.
DR SMART: SM00291; Znf_ZZ; 1.
DR PROSITE: PS01357; ZF_ZZ_1; 1.
DR PROSITE: PS0135; ZF_ZZ_2; 1.
KW Coiled coil; Zinc-finger; Alternative splicing.
FT ZN_FING 237 284
FT DOMAIN 269 418
FT DOMAIN 428 521
FT VARSPLIC 518 518
FT VARSPLIC 603 608
FT VARSPLIC 609 700
FT CONFLICT 412 412
FT CONFLICT 465 465
SQ SEQUENCE 700 AA; 78355 MW; 6BE34EEB61AFE46C CRC64;

Query Match      51.9%; Score 42; DB 1; Length 700;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSGAGIAQAAAG 13
Db ||| | : |||
653 GSGEGLPEAAAG 664

```

Search completed: March 10, 2004, 12:06:30
 Job time : 6.23077 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:57:36 ; Search time 26.9167 Seconds
(without alignments)
199.275 Million cell updates/sec

Title: US-10-044-703-49

Perfect score: 81

Sequence: 1 TGSAGIAQAAGTVNI 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mic.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81	100.0	374	16 Q7U0Z8	Q7U0Z8 mycobacteri
2	61	75.3	374	2 Q49590	Q49590 mycobacteri
3	60	74.1	336	2 Q49675	Q49675 mycobacteri
4	60	74.1	374	2 Q49589	Q49589 mycobacteri
5	58	71.6	300	16 Q88YM2	Q88YM2 lactobacill
6	55	67.9	295	16 Q895Y3	Q895Y3 clostridriu
7	54	66.7	338	17 Q8TND9	Q8TND9 methanosarc
8	53	65.4	363	16 Q8PM56	Q8PM56 xanthomonas
9	53	65.4	363	16 Q8PAF7	Q8PAF7 xanthomonas
10	51	63.0	140	2 Q66079	Q66079 lactococcus
11	51	63.0	283	16 Q9CEW4	Q9CEW4 lactococcus
12	50	61.7	1598	5 Q95YM8	Q95YM8 apis mellif
13	49	60.5	347	16 Q8YVD9	Q8YVD9 anabaena sp
14	48	59.3	299	16 Q8IUZ1	Q8IUZ1 bacillus an
15	48	59.3	309	16 Q8IHS2	Q8IHS2 bacillus ce
16	48	59.3	492	2 Q9FAE8	Q9FAE8 acidovorax

17	47	58.0	298	16 Q9CEW3	Q9CEW3 lactococcus
18	47	58.0	341	16 Q880A3	Q880A3 pseudomonas
19	47	58.0	476	2 Q86457	Q86457 rhodobacter
20	46	56.8	239	16 Q8NRC1	Q8NRC1 corynebacte
21	46	56.8	296	16 Q834F0	Q834F0 enterococcu
22	46	56.8	435	5 Q9VLA6	Q9VLA6 drosophila
23	46	56.8	492	2 Q9FAE7	Q9FAE7 acidovorax
24	46	56.8	966	10 Q84UT7	Q84UT7 oryza sativ
25	46	56.8	3010	12 Q68285	Q68285 hepatitis c
26	46	56.8	3010	12 Q81825	Q81825 hepatitis c
27	45	55.6	35	2 Q94S7	Q94S7 rhodobacter
28	45	55.6	145	2 Q93AL4	Q93AL4 nodularia s
29	45	55.6	245	16 Q8EM35	Q8EM35 shewanella
30	45	55.6	255	16 Q8FZ74	Q8FZ74 brucella su
31	45	55.6	259	16 Q7WIZ9	Q7WIZ9 bordetella
32	45	55.6	259	16 Q7W6W0	Q7W6W0 bordetella
33	45	55.6	264	2 Q85723	Q85723 streptomyce
34	45	55.6	284	16 Q834B0	Q834B0 enterococcu
35	45	55.6	292	16 Q97031	Q97031 streptococc
36	45	55.6	292	16 Q8DPB1	Q8DPB1 streptococc
37	45	55.6	304	16 Q8YIN7	Q8YIN7 brucella me
38	45	55.6	387	2 Q83TF1	Q83TF1 pseudomonas
39	45	55.6	392	16 Q8YIN7	Q8YIN7 anabaena sp
40	45	55.6	544	16 Q8F013	Q8F013 leptospira
41	45	55.6	3010	12 Q68533	Q68533 hepatitis c
42	44	54.3	203	5 Q8IQ26	Q8IQ26 drosophila
43	44	54.3	247	16 Q81KK8	Q81KK8 bacillus an
44	44	54.3	265	2 Q9KI91	Q9KI91 bacillus an
45	44	54.3	286	16 Q8DZV4	Q8DZV4 streptococc

ALIGNMENTS

RESULT 1

Q7U0Z8 PRELIMINARY; PRT; 374 AA.
 AC Q7U0Z8;
 DT 01-OCT-2003 (Tremblrel. 25, Created)
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Periplasmic phosphate-binding lipoprotein PSTS1 (PBP-1) (PST1).
 GN PSTS1 OR MB0959.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709197; PubMed=12788972;
 RA Garnier T., Bigmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248337; CAD93820.1; -.
 KW Complete proteome.
 SQ SEQUENCE 374 AA; 38215 MW; 1434968191PF201D CRC64;

Query Match 100.0%; Score 81; DB 16; Length 374;
 Best Local Similarity 100.0%; Pred. No. 0.0012;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSAGIAQAAGTVNI 17

Db 86 TGSAGIAQAAGTVNI 102

RESULT 2

Q49590 PRELIMINARY; PRT; 374 AA.
 ID Q49590

```

AC Q49590;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PstS subunit of ABC transporter.
GN PSTS1B.
OS Mycobacterium intracellulare.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1767;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35761;
RX MEDLINE=97055782; PubMed=8900068;
RA Thangaraj H.S., Bull T.J., De Smet K.A.L., Hill M., Rouse D.A.,
RA Moreno C., Ivanyi J.;
RT "Duplication of genes encoding the immunodominant 38 kDa antigen in
RT FEMS Microbiol. Lett. 144:235-240 (1996).
RL EMBL; X95538; I1XG.
DR HSP; P06128; 11XG.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000437; Prok_lipoprot_S.
DR Pfam; PF01547; SBP_bac_1; 1.
DR ProSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
SQ SEQUENCE 374 AA; 38088 MW; A7F2D5843860B9D7 CRC64;

Query Match 75.3%; Score 61; DB 2; Length 374;
Best Local Similarity 76.5%; Pred. No. 0.91; Mismatches 1; Indels 0; Gaps 0;
Matches 13; Conservative 1;

QY 1 TQSGAGIAQAAGTVNI 17
DB 85 TQSGTGISQAAGAVAI 101

RESULT 3
Q49675
ID Q49675 PRELIMINARY; PRT; 336 AA.
AC Q49675;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE PabB.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96084954; PubMed=7489918;
RA Doukhan L., Predlich M., Nair G., Duseurget O., Mandic-Mulec I.,
RA Cole S.T., Smith D.R., Smith I.;
RT "Genomic organization of the mycobacterial sigma gene cluster.";
RL Gene 165:67-70(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U00012; AAAS5925.1; -.
SQ SEQUENCE 336 AA; 35916 MW; 0CE7CABDB3CBD99F CRC64;

Query Match 74.1%; Score 60; DB 2; Length 336;
Best Local Similarity 76.5%; Pred. No. 1.1; Mismatches 1; Indels 0; Gaps 0;
Matches 13; Conservative 1;

QY 1 TQSGAGIAQAAGTVNI 17
DB 80 TQGTGIAQAAGTVDI 96

```

```

RESULT 4
Q49589
ID Q49589 PRELIMINARY; PRT; 374 AA.
AC Q49589;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PstS subunit of ABC transporter.
GN PSTS1A.
OS Mycobacterium intracellulare.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1767;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35761;
RX MEDLINE=97055782; PubMed=8900068;
RA Thangaraj H.S., Bull T.J., De Smet K.A.L., Hill M., Rouse D.A.,
RA Moreno C., Ivanyi J.;
RT "Duplication of genes encoding the immunodominant 38 kDa antigen in
RT FEMS Microbiol. Lett. 144:235-240 (1996).
RL EMBL; X95538; CA64783.1; -.
DR HSP; P06128; IAS4.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000437; Prok_lipoprot_S.
DR InterPro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1; 1.
DR ProSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
SQ SEQUENCE 374 AA; 37998 MW; 45796D4E9F6F513D CRC64;

Query Match 74.1%; Score 60; DB 2; Length 374;
Best Local Similarity 76.5%; Pred. No. 1.3; Mismatches 2; Indels 0; Gaps 0;
Matches 13; Conservative 2;

QY 1 TQSGAGIAQAAGTVNI 17
DB 86 TQSGAGISQAAGAVVI 102

RESULT 5
Q88YM2
ID Q88YM2 PRELIMINARY; PRT; 300 AA.
AC Q88YM2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phosphate ABC transporter, substrate binding protein.
GN PSTF OR LP 0733.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OX Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=22480296; PubMed=12566566;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.B.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hoffer S.M., Nierop Groot M.N., Kerckhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
DR EMBL; AL935254; CAD33329.1; -.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000437; Prok_lipoprot_S.
DR InterPro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1; 1.
DR ProSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR Complete proteome.
KW

```

SQ SEQUENCE 300 AA; 31480 MW; FBB7D52A83355E45 CRC64;
 Query Match 71.6%; Score 58; DB 16; Length 300;
 Best Local Similarity 68.8%; Pred. No. 2;
 Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

 QY 2 GSGAGTAAAGTNNI 17
 |||||:|:|:|
 Db 78 GSGAGLSQVAKGSNNI 93

 RESULT 6
 Q895Y3 PRELIMINARY; PRT; 295 AA.
 AC Q895Y3
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Phosphate-binding periplasmic protein.
 GN C7C01132
 OS Clostridium tetani.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Massachusetts / E98;
 RX MEDLINE=22457253; PubMed=12552129;
 RA Brueggemann H., Baumer S., Fricke W.F., Wier A., Liesegang H.,
 RA Decker I., Herberg C., Martinez-Arias R., Merkl R., Henne A.,
 RA Gottschalk G.;
 RT "The genome sequence of Clostridium tetani, the causative agent of
 RT tetanus disease.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321 (2003).
 DR ENBL; A015939; AAC35707.1; -.
 DR GO; GO:0005215; P:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000437; Prok_lipprot_S.
 DR InterPro; IPR006059; SBP_bac_1.
 DR Pfam; PF01547; SBP_bac_1; 1.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Complete proteome.
 SQ SEQUENCE 295 AA; 32385 MW; 4E875E845C317388 CRC64;

 Query Match 67.9%; Score 55; DB 16; Length 295;
 Best Local Similarity 68.8%; Pred. No. 5.2;
 Matches 11; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

 QY 2 GSGAGTAAAGTNNI 17
 |||||:|:|:|
 Db 79 GSGAGISQVSEGTVEI 94

 RESULT 7
 Q8TND9 PRELIMINARY; PRT; 338 AA.
 AC Q8TND9
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Phosphate-binding protein.
 GN MA2349.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11922238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,

RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Unayam D.A., White O., White R.H., de Macario E.C.,
 RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Fritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity.";
 RL Genome Res. 12:532-542 (2002).
 DR ENBL; A010923; AAM05739.1; -.
 KW Complete proteome.
 SQ SEQUENCE 338 AA; 36061 MW; A6AEAC150F66EEES CRC64;

 Query Match 66.7%; Score 54; DB 17; Length 338;
 Best Local Similarity 68.8%; Pred. No. 8.4;
 Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

 QY 2 GSGAGTAAAGTNNI 17
 |||||:|:|:|
 Db 103 GSGAGISSAGTGVNI 118

 RESULT 8
 Q8PM56 PRELIMINARY; PRT; 363 AA.
 AC Q8PM56
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE ABC transporter phosphate binding protein.
 GN PSTS OR XA1577
 OS Xanthomonas axonopodis (pv. citri).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=92829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=2202145; PubMed=1204217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quagga R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Canarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.P.,
 RA Setubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities.";
 RL Nature 417:459-463 (2002).
 DR ENBL; A011790; AAM36445.1; -.
 DR GO; GO:0005215; P:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000437; Prok_lipprot_S.
 DR InterPro; IPR006059; SBP_bac_1.
 DR Pfam; PF01547; SBP_bac_1; 1.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Complete proteome.
 SQ SEQUENCE 363 AA; 37641 MW; 1A39C0289BF7DBDF CRC64;

 Query Match 65.4%; Score 53; DB 16; Length 363;
 Best Local Similarity 73.3%; Pred. No. 13;
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mblk-1 protein.
GN MBLK-1.

OS Apis mellifera (Honeybee).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
OC Apidae; Apis.
OX NCBI_TaxID=7460;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=21873258; PubMed=11881813;
RA Takeuchi H., Kage E., Sawata M., Kamikouchi A., Ohashi K., Ohara M.,
RA Fujiyuki T., Kunieda T., Sekimizu K., Natori S., Kubo T.,
RT "Identification of a novel gene, Mblk-1, that encodes a putative
RT transcritption factor expressed preferentially in the large-type Kenyon
RL cells of the honey bee brain."
RL Insect Mol. Biol. 10:487-494(2001).

DR EMBL; AB047034; BAB64310.1; -.
DR InterPro; IPR007889; HTH_psq.
DR Pfam; PF05225; HTH_psq.
DR NCBI_TaxID=7460;
SQ SEQUENCE 1598 AA; 174329 MW; E5475BDD3ACB1EF CRC64;

Query Match 61.7%; Score 50; DB 5; Length 1598;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 GSGAGIAQAAGTGN 16

Db 1241 GVGAGIAETSGTGN 1255

RESULT 13

Q8YYD9 PRELIMINARY; PRT; 347 AA.

DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE ABC phosphate transport system phosphate-binding periplasmic
DE protein.
GN AL0911.

OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.

OX NCBI_TaxID=103690;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).

DR EMBL; AP003564; BAB72868.1; -.
DR PIR; AD1920; AD1920.

DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR006059; SBP_bac_1.

DR Pfam; PF01547; SBP_bac_1; 1.

KW Complete proteome.
SQ SEQUENCE 347 AA; 37432 MW; 450712646359098C CRC64;

Query Match 60.5%; Score 49; DB 16; Length 347;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 GSGAGIAQAAGTGN 16

Db 70 GSGGGINQTITGTV 84

RESULT 14

Q81U21 PRELIMINARY; PRT; 299 AA.

AC Q81U21;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phosphate ABC transporter, phosphate-binding protein, putative.
GN BA0715.

OS Bacillus anthracis (strain Ames).

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=158094;
RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.D., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria."
RL Nature 423:81-86(2003).

DR EMBL; AE017026; AAP24726.1; -.

DR TIGR; BA0715; -.

DR InterPro; IPR000437; Prok_lipoprot S.

DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.

KW Complete proteome.

SQ SEQUENCE 299 AA; 31553 MW; 44C0D3613D2938C5 CRC64;

Query Match 59.3%; Score 48; DB 16; Length 299;
Best Local Similarity 62.5%; Pred. No. 55;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 GSGAGIAQAAGTGN 17

Db 82 GSGTGINQVAGVQI 97

RESULT 15

Q81HS2 PRELIMINARY; PRT; 309 AA.

AC Q81HS2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Phosphate-binding protein.

GN BC0710.

OS Bacillus cereus (strain ATCC 14579 / DSM 31).

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=226900;
RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapatal V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Vazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Fusch G., Hasekorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis."
RL Nature 423:87-91(2003).

DR EMBL; AE017000; AAP07723.1; -.

KW Complete proteome.

SQ SEQUENCE 309 AA; 32745 MW; CD2F0EEFCB91E519 CRC64;

Query Match 59.3%; Score 48; DB 16; Length 309;

Best Local Similarity 62.5%; Pred. No. 56;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 GSGAGIAQAAAGTVNI 17
Db 92 GSGTGINQVAGAVOI 107

Search completed: March 10, 2004, 12:10:51
Job time : 28.0417 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:50:51 ; Search time 47.5641 Seconds
(without alignments)
118.807 Million cell updates/sec

Title: US-10-044-703-52

Perfect score: 103

Sequence: 1 AETPGCVAYIGISFLDQASQ 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	100.0	20	AAE12267	Aae12267 Mycobacte
2	103	100.0	351	ABU63550	Abu63550 Mature My
3	103	100.0	373	ABU63549	Abu63549 Mycobacte
4	103	100.0	374	AAR30090	Aar30090 38 kDa li
5	103	100.0	374	AAW64363	Aaw64363 Mycobacte
6	103	100.0	374	AAW64364	Aaw64364 Mycobacte
7	103	100.0	374	AAW81730	Aaw81730 M. tuberc
8	103	100.0	374	AAW81731	Aaw81731 M. tuberc
9	103	100.0	374	AAW32061	Aay32061 Mycobacte
10	103	100.0	374	AAW39018	Aay39018 M. tuberc
11	103	100.0	374	AAW39017	Aay39017 M. tuberc
12	103	100.0	374	AAW39161	Aay39161 M. tuberc
13	103	100.0	374	AAW39160	Aay39160 M. tuberc
14	103	100.0	374	AAE11840	Aae11840 Mycobacte
15	103	100.0	374	AAE29719	Aae29719 Mycobacte
16	103	100.0	374	AAW50733	Aaw50733 Mycobacte
17	103	100.0	374	AAE17593	Aae17593 Mycobacte
18	103	100.0	374	AAU74590	Aau74590 Mycobacte
19	103	100.0	652	AAW39082	Aaw39082 M. tuberc
20	103	100.0	652	AAW39225	Aaw39225 M. tuberc
21	103	100.0	802	AAW64379	Aaw64379 Mycobacte
22	103	100.0	802	AAW81746	Aaw81746 M. tuberc
23	103	100.0	802	AAW32063	Aay32063 Mycobacte
24	103	100.0	802	AAW39033	Aay39033 M. tuberc
25	103	100.0	802	AAW39081	Aay39081 M. tuberc

ALIGNMENTS

RESULT 1

AAE12267
ID AAE12267 standard; peptide; 20 AA.

XX AC AAE12267;

XX DT 18-DEC-2001 (first entry)

XX DE Mycobacterium tuberculosis (Mtb) peptide #52.

XX KW Mycobacterium tuberculosis; Mtb peptide; antibacterial; vaccine;

XX KW infection; anti-Mtb immune response.

XX OS Mycobacterium tuberculosis.

XX PN WO200170774-A2.

XX PD 27-SEP-2001.

XX PF 20-MAR-2001; 2001WO-US008906.

XX PR 20-MAR-2000; 2000US-0190834P.

XX PA (UYBR-) UNIV BROWN RES FOUND.

XX PI Degroot AS;

XX DR WPI; 2001-616401/71.

XX PT New vaccine for immunizing a mammalian subject, preferably humans,

XX PT against infection caused by Mycobacterium tuberculosis.

XX PS Disclosure; Fig 4; 42pp; English.

XX CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti- Mycobacterium tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a mammalian subject preferably human. They are used for immunising a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

SQ Sequence 20 AA;

Query Match 100.0%; Score 103; DB 4; Length 20;

Aay39176 M. tuberc
Aay39224 M. tuberc
Aau74592 Antigenic
Aau01901 M. tuberc
Aac95326 E. faeciu
Abb49989 Listeria
Abb55070 Lactococc
Abb26440 Streptoco
Aau41131 Propionib
Abm37650 Propionib
Abm37650 Propionib
Abp30084 Streptoco
Abp26260 Streptoco
Adc94775 E. faeciu
Aau36131 Klebsiell
Abb71568 Drosophil
Aag42167 Arabidops
Aag45436 Arabidops
Aag11561 Arabidops
Aag11560 Arabidops
Aag45435 Arabidops

26 103 100.0 802 2 AAY39176
27 103 100.0 802 2 AAY39224
28 103 100.0 802 5 AAU74592
29 103 100.0 983 4 AAU01901
30 58 56.3 301 7 ADC95326
31 52 50.5 302 5 ABB49989
32 50 48.5 298 5 ABB55070
33 47 45.6 189 5 ABP26440
34 47 45.6 256 4 AAU41131
35 47 45.6 256 6 ABM37650
36 47 45.6 286 5 ABP30084
37 46 44.7 288 5 ABP26260
38 46 44.7 289 7 ADC94775
39 46 44.7 395 4 AAU36131
40 46 44.7 844 4 ABB71568
41 45 43.7 311 3 AAG42167
42 45 43.7 311 3 AAG45436
43 45 43.7 311 3 AAG11561
44 45 43.7 464 3 AAG11560
45 45 43.7 464 3 AAG45435

Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
|||||
Db 1 AETPGCVAYIGISFLDQASQ 20

RESULT 2

ABU63550
ID ABU63550 standard; protein; 351 AA.

XX AC ABU63550;
XX

DT 25-SEP-2003 (first entry)

XX DE Mature Mycobacterium tuberculosis 38kDa protein.

XX KW Antimicrobial; cytostatic; type 1 response; 38kDa protein;
XX KW T-helper-1 T-cell response; interleukin-12; IL-12;
XX KW cell-mediated immunity; pathogen infection; vaccine;
XX KW interferon-gamma sensitive tumour.

XX OS Mycobacterium tuberculosis.

XX FH Key Location/Qualifiers
XX FT Modified-site 1
XX FT /label= OTHER
XX FT /note= "OTHER= N-acyl diglyceride cysteine"

XX PN US6517839-B1.
XX PD 11-FEB-2003.
XX PF 17-JUL-1998; 98US-00118426.
XX PR 18-JUL-1997; 97US-0052970P.
XX PS (REGC) UNIV CALIFORNIA.

XX PI Modlin RL, Libraty DH;
XX DR WPI; 2003-553609/52.

XX CC The invention describes a method of inducing type 1/T-helper-1 T-cell response to
XX CC lipopeptide (I) having an N-terminal ester- or amide-linked fatty acyl
XX CC group. The methods are used to stimulate cell-mediated immunity for
XX CC prevention and treatment of pathogen infections (e.g. Mycobacterium
XX CC tuberculosis) and for treatment of a interferon-gamma sensitive tumour.
XX CC Also useful in the development of new vaccine and therapeutic strategies.
XX CC This is the amino acid sequence of mature Mycobacterium tuberculosis
XX CC 38kDa protein from which lipopeptides can be isolated

XX SQ Sequence 351 AA;

Query Match 100.0%; Score 103; DB 6; Length 351;
Best Local Similarity 100.0%; Pred. No. 2.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
|||||
Db 216 AETPGCVAYIGISFLDQASQ 235

RESULT 3

Query Match 100.0%; Score 103; DB 6; Length 351;
Best Local Similarity 100.0%; Pred. No. 2.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
|||||
Db 216 AETPGCVAYIGISFLDQASQ 235

Query Match 100.0%; Score 103; DB 6; Length 351;
Best Local Similarity 100.0%; Pred. No. 2.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
|||||
Db 216 AETPGCVAYIGISFLDQASQ 235

RESULT 4

Query Match 100.0%; Score 103; DB 6; Length 351;
Best Local Similarity 100.0%; Pred. No. 2.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

Query Match 100.0%; Score 103; DB 6; Length 351;
Best Local Similarity 100.0%; Pred. No. 2.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 6

Query Match 100.0%; Score 103; DB 6; Length 351;
Best Local Similarity 100.0%; Pred. No. 2.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ABU63549
ID ABU63549 standard; protein; 373 AA.

XX AC ABU63549;
XX

DT 25-SEP-2003 (first entry)

XX DE Mycobacterium tuberculosis 38kDa protein.

XX KW Antimicrobial; cytostatic; type 1 response; 38kDa protein;
XX KW T-helper-1 T-cell response; interleukin-12; IL-12;
XX KW cell-mediated immunity; pathogen infection; vaccine;
XX KW interferon-gamma sensitive tumour.

XX OS Mycobacterium tuberculosis.

XX FH Key Location/Qualifiers
XX FT Modified-site 1
XX FT /label= OTHER
XX FT /note= "OTHER= N-acyl diglyceride cysteine"

XX PN US6517839-B1.

XX PD 11-FEB-2003.

XX PF 17-JUL-1998; 98US-00118426.

XX PR 18-JUL-1997; 97US-0052970P.

XX PS (REGC) UNIV CALIFORNIA.

XX PI Modlin RL, Libraty DH;
XX DR WPI; 2003-553609/52.

XX CC The invention describes a method of inducing type 1/T-helper-1 T-cell response to
XX CC lipopeptide (I) having an N-terminal ester- or amide-linked fatty acyl
XX CC group. The methods are used to stimulate cell-mediated immunity for
XX CC prevention and treatment of pathogen infections (e.g. Mycobacterium
XX CC tuberculosis) and for treatment of a interferon-gamma sensitive tumour.
XX CC Also useful in the development of new vaccine and therapeutic strategies.
XX CC This is the amino acid sequence of Mycobacterium tuberculosis 38kDa
XX CC protein from which lipopeptides can be isolated

XX SQ Sequence 373 AA;

Query Match 100.0%; Score 103; DB 6; Length 373;
Best Local Similarity 100.0%; Pred. No. 3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
|||||
Db 238 AETPGCVAYIGISFLDQASQ 257

XX Tubercule bacilli protein; tuberculosis; diagnosis; TB; peptide.
 XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers
 XX Peptide 1..20
 XX Peptide /note= "peptide A"
 XX Peptide 24..43
 XX Region /note= "peptide H"
 XX Peptide 66..76
 XX Region /note= "amphipathic region"
 XX Peptide 146..165
 XX Region /note= "peptide B"
 XX Peptide 150..164
 XX Region /note= "amphipathic region"
 XX Peptide 201..220
 XX Region /note= "peptide C"
 XX Peptide 204..210
 XX Region /note= "amphipathic region"
 XX Peptide 210..229
 XX Region /note= "peptide D"
 XX Peptide 216..225
 XX Region /note= "amphipathic region"
 XX Peptide 230..249
 XX Region /note= "peptide E"
 XX Peptide 236..243
 XX Region /note= "amphipathic region"
 XX Peptide 285..304
 XX Region /note= "peptide F"
 XX Peptide 287..291
 XX Region /note= "amphipathic region"
 XX Peptide 296..304
 XX Region /note= "amphipathic region"
 XX Peptide 350..369
 XX Region /note= "peptide G"
 XX Peptide 362..368
 XX Region /note= "amphipathic region"

XX W09221697-A2.

XX 10-DEC-1992.

XX 26-MAY-1992; 92WO-GB000948.

XX 24-MAY-1991; 91GB-00011291.

XX (MEDI-) MEDICAL RES COUNCIL.

XX Vordermeier H, Harris D, Moreno C, Ivanyi J;

XX WPI; 1992-433610/52.

XX Peptide(s) from mycobacterium tuberculosis antigens - useful for the
 XX diagnosis of tuberculosis.

XX Disclosure; Fig 1; 44pp; English.

XX The sequence is that of the 38 kDa lipoprotein antigen of Mycobacterium
 XX tuberculosis, from this peptides can be derived which can be used in
 XX place of the purified protein derivative (PPD) test. The peptides can be
 XX used to diagnose tuberculosis (TB) in a human or non-human animal, and to
 XX distinguish a TB patient from an infected or otherwise sensitised but
 XX healthy clinical suspect. They stimulate lymphocytes which have been
 XX sensitised to antigens of M. tuberculosis. The peptides may be used in
 XX vivo skin tests relying on delayed hypersensitivity reaction causing
 XX reddening and swelling, or in ex vivo detection of activated lymphocytes.
 XX They have excellent sensitivity and improved specificity to PPD-
 XX tuberculin. They fail to provoke a strong immune response in delayed-type
 XX hypersensitivity or lymphocyte activation tests in patients with
 XX pulmonary and non-lymphatic extrapulmonary TB. This provides the basis of
 XX distinguishing TB patients from these patients. Diagnosing TB comprises
 XX either intradermal injection of the peptide, or contacting lymphocytes

CC with the peptide. (Updated on 25-MAR-2003 to correct PN field.)
 XX
 XX Sequence 374 AA;

Query Match 100.0%; Score 103; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
 |||||
 DB 239 AETPGCVAYIGISFLDQASQ 258
 |||||

RESULT 5

AAW64363

ID AAW64363 standard; protein; 374 AA.

XX AAW64363;

XX 09-NOV-1998 (first entry)

XX Mycobacterium antigen TbRa3-38 kD-Tb38-1 fusion protein.

XX Tuberculosis; infection; diagnosis; 38 kDa antigen; TbRa3; Tb38-1.

XX Mycobacterium tuberculosis.

XX Synthetic.

XX W09816645-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US018214.

XX 11-OCT-1996; 96US-00729622.

XX 13-MAR-1997; 97US-00818111.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

XX Vedvick TS, Twardzik DR, Lodes MJ;

XX WPI; 1998-251292/22.

XX N-PSDB; AAV44413.

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
 XX develop products for the detection of M. tuberculosis infection and
 XX diagnosis of tuberculosis.

XX Example 7; Page 165-166; 250pp; English.

XX This polypeptide comprises a fusion protein composed of Mycobacterium
 XX tuberculosis antigens TbRa3 (see AAW64295), 38 kDa antigen (see AAW64364)
 XX and Tb38-1 (see AAW64321). It was produced by PCR amplification (see
 XX AAV4407-12) of the appropriate antigen DNA sequences (see AAV44413) and
 XX cloning into an expression vector. The fusion protein was expressed in E.
 XX coli. It can be used for serodiagnosis of tuberculosis. The invention
 XX relates to compositions and methods for diagnosing tuberculosis. It
 XX provides polypeptides (see AAW64291-W64379) comprising an antigenic
 XX portion of a soluble M. tuberculosis antigen, an immunogenic portion of
 XX an M. tuberculosis antigen, or a fusion protein, as well as DNA sequences
 XX encoding such polypeptides, recombinant expression vectors and
 XX transformed or transfected host cells. Also claimed are methods and
 XX diagnostic kits for detecting M. tuberculosis infection in a patient

XX Sequence 374 AA;

Query Match 100.0%; Score 103; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
 |||||

Db 239 AETPGCVAYIGISFLDQASQ 258

RESULT 6

AAW64364
ID AAW64364 standard; protein; 374 AA.

AC AAW64364;
XX

DT 09-NOV-1998 (first entry)
XX

DE Mycobacterium 39 kDa antigen.
XX

KW Tuberculosis; infection; diagnosis; 38 kDa antigen.
XX

OS Mycobacterium tuberculosis.
XX

PN WO9816645-A2.
XX

PD 23-APR-1998.
XX

PF 07-OCT-1997; 97WO-US018214.
XX

PR 11-OCT-1996; 96US-00729622.
XX

PR 13-MAR-1997; 97US-00818111.
XX

PA (CORI-) CORIXA CORP.
XX

PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX

PI Vedvick TS, Twardzik DR, Lodes MJ;
XX

DR WPI; 1998-251292/22.
XX

DR N-PSDB; AAV44414.
XX

PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
PT develop products for the detection of M. tuberculosis infection and
PT diagnosis of tuberculosis.
XX

PS Disclosure; Page 168-169; 250pp; English.
XX

CC This polypeptide comprises Mycobacterium tuberculosis 38 kDa antigen. A
CC Tba3-38 KD-Tb38-1 fusion protein (see AAW64363) has been constructed
CC that can be used in the serodiagnosis of tuberculosis. The invention
CC relates to compositions and methods for diagnosing tuberculosis. It
CC provides polypeptides (see AAW64291-W64379) comprising an antigenic
CC portion of a soluble M. tuberculosis antigen, an immunogenic portion of
CC an M. tuberculosis antigen, or a fusion protein, as well as DNA sequences
CC encoding such polypeptides, recombinant expression vectors and
CC transformed or transfected host cells. Also claimed are methods and
CC diagnostic kits for detecting M. tuberculosis infection in a patient
XX

SQ Sequence 374 AA;
XX

Query Match 100.0%; Score 103; DB 2; Length 374;
Best Local Similarity 100.0%; Pred. No. 3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
|||||

Db 239 AETPGCVAYIGISFLDQASQ 258
|||||

RESULT 7

AAW81730
ID AAW81730 standard; protein; 374 AA.

AC AAW81730;
XX

DT 27-JAN-1999 (first entry)
XX

DE M. tuberculosis fusion protein Tba3/38kD/Tb38-1.
XX

KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
XX

KW vaccine; pharmaceutical; infection; diagnosis.
XX

OS Synthetic.
OS

OS Mycobacterium tuberculosis.
XX

PN WO9816646-A2.
XX

PD 23-APR-1998.
XX

PF 07-OCT-1997; 97WO-US018293.
XX

PR 11-OCT-1996; 96US-00730510.
XX

PR 13-MAR-1997; 97US-00818112.
XX

PA (CORI-) CORIXA CORP.
XX

PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ;
XX

DR WPI; 1998-261042/23.
XX

DR N-PSDB; AAV64522.
XX

CC Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
CC develop products for the detection of M. tuberculosis infection and for
CC diagnosis, treatment and prevention of tuberculosis.
XX

PS Claim 37; Page 156-157; 230pp; English.
XX

CC This sequence represents a fusion protein containing the immunogenic
CC polypeptides Tba3, 38kD and Tb38-1 from Mycobacterium tuberculosis (MT).
CC This fusion protein can be used in a method for inducing protective
CC immunity against tuberculosis (TB). This sequence can be formulated into
CC vaccines and/or pharmaceutical compositions for immunising against M.
CC tuberculosis infection or may be used for the diagnosis of TB
XX

SQ Sequence 374 AA;
XX

Query Match 100.0%; Score 103; DB 2; Length 374;
Best Local Similarity 100.0%; Pred. No. 3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
|||||

Db 239 AETPGCVAYIGISFLDQASQ 258
|||||

RESULT 8

AAW81731

ID AAW81731 standard; protein; 374 AA.

AC AAW81731;
XX

DT 27-JAN-1999 (first entry)
XX

DE M. tuberculosis 38kD antigen protein.
XX

KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX

OS Mycobacterium tuberculosis.
XX

PN WO9816646-A2.
XX

PD 23-APR-1998.
XX

PF 07-OCT-1997; 97WO-US018293.
XX

PR 11-OCT-1996; 96US-00730510.
XX

PR 13-MAR-1997; 97US-00818112.
XX

PA (CORI-) CORIXA CORP.
XX

PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX

PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX WPI; 1998-261042/23.
 XX N-PSDB; AAV64523.
 XX
 PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and for
 PT diagnosis, treatment and prevention of tuberculosis.
 XX
 PS Claim 23; Page 159-160; 230pp; English.
 XX
 CC This sequence represents a 38kD antigen from Mycobacterium tuberculosis
 CC (MT). This antigen can be used in a method for inducing protective
 CC immunity against tuberculosis (TB). This sequence can be formulated into
 CC vaccines and/or pharmaceutical compositions for immunising against M.
 CC tuberculosis infection or may be used for the diagnosis of TB
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 103; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AETPGCVAYIGISFLDQASQ 20
 |||||
 DB 239 AETPGCVAYIGISFLDQASQ 258
 |||||
 RESULT 9
 AAY32061
 ID AAY32061 standard; protein; 374 AA.
 AC AAY32061;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen Tbra3-38kD-Tb38-1 fusion.
 XX
 KW Tuberculosis; antigen; fusion protein; Tbra3; 38kD; Tb38-1; diagnosis;
 KW therapy; vaccine; immunogen.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN W09951748-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 07-APR-1999; 99WO-US007717.
 XX
 PR 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky YAW, Alderson M, Campos-Neto A;
 XX
 DR WPI; 1999-601610/51.
 DR N-PSDB; AAZ20196.
 XX
 PT New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis.
 XX
 PS Claim 1; Fig 3A-D; 83pp; English.
 XX
 CC This sequence represents a recombinant Mycobacterium tuberculosis tri-
 CC antigen fusion protein composed of the antigens Tbra3, 39kD and Tb38-1.
 CC The fusion protein is expressed in host cells using a vector carrying a
 CC polynucleotide (see AAZ20196) comprising the 3 coding sequences for the
 CC antigens. The invention provides fusion proteins (see AAY32059-71)
 CC containing at least 2 M. tuberculosis antigens. The new fusion proteins
 CC and polynucleotides encoding them are useful as vaccines for preventing
 CC tuberculosis (claimed), for diagnosis (via in vitro assays or intradermal
 CC skin tests for detection of anti-M. tuberculosis antibodies), monitoring

CC of disease progression, and treatment of tuberculosis. They are more
 CC effective immunogens than mixtures of the individual protein components
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 103; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AETPGCVAYIGISFLDQASQ 20
 |||||
 DB 239 AETPGCVAYIGISFLDQASQ 258
 |||||
 RESULT 10
 AAY39018
 ID AAY39018 standard; protein; 374 AA.
 XX
 AC AAY39018;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis 38 kDa antigen protein.
 XX
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN W09942118-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US003265.
 XX
 PR 18-FEB-1998; 98US-00024753.
 PR 05-MAY-1998; 98US-00072596.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX
 DR WPI; 1999-527416/44.
 DR N-PSDB; AAZ19112.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis.
 PS Claim 53; Page 199-200; 323pp; English.
 XX
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against M.
 CC tuberculosis infection. The new detection methods are needed as current
 CC vaccination strategies do not provide 100% immunity
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 103; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AETPGCVAYIGISFLDQASQ 20
 |||||
 DB 239 AETPGCVAYIGISFLDQASQ 258
 |||||
 RESULT 11
 AAY39017
 ID AAY39017 standard; protein; 374 AA.
 XX

AC AAY39017;
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis fusion protein TbrA3/38kDa/Tb38-1.
 XX
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942118-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US003265.
 XX
 PR 18-FEB-1998; 98US-00024753.
 PR 05-MAY-1998; 98US-00072596.
 XX
 PA (CORI-) CORIXA CORP.
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX
 DR WPI; 1999-527416/44.
 DR N-PSDB; AA219111.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis.
 XX
 PS Example 10; Page 196-198; 323pp; English.
 XX
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against M.
 CC tuberculosis infection. The new detection methods are needed as current
 CC vaccination strategies do not provide 100% immunity
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 103; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AETPGCVAYIGISFLDQASQ 20
 DB 239 AETPGCVAYIGISFLDQASQ 258
 RESULT 12
 AAY39161
 ID AAY39161 standard; protein; 374 AA.
 XX
 AC AAY39161;
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis antigen 38 kD amino acid sequence.
 XX
 DE
 KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942076-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US003268.
 XX

PR 18-FEB-1998; 98US-00025197.
 PR 05-MAY-1998; 98US-00072967.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX
 DR WPI; 1999-527409/44.
 DR N-PSDB; AA219324.
 XX
 XX New antigens from Mycobacterium tuberculosis useful in diagnostic skin
 PT tests and protective or therapeutic vaccines or compositions.
 XX
 PS Claim 23; Page 154-155; 299pp; English.
 XX
 CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
 CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
 CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
 CC polypeptides fragments, can be used in pharmaceutical compositions or
 CC vaccines to generate a protective or therapeutic immune response to M.
 CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
 CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
 CC killer cells and/or macrophages in tuberculosis-immune subjects. AA219249
 CC to AA219460 and AAY39083 to AAY39225 are used in the exemplification of
 CC the present invention
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 103; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AETPGCVAYIGISFLDQASQ 20
 DB 239 AETPGCVAYIGISFLDQASQ 258
 RESULT 13
 AAY39160
 ID AAY39160 standard; protein; 374 AA.
 XX
 AC AAY39160;
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis TbrA3, 38 kD and Tb38-1 fusion protein.
 XX
 KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.
 XX
 OS Synthetic.
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942076-A2.
 XX
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US003268.
 XX
 PR 18-FEB-1998; 98US-00025197.
 PR 05-MAY-1998; 98US-00072967.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX
 DR WPI; 1999-527409/44.
 DR N-PSDB; AA219323.
 XX

PT New antigens from Mycobacterium tuberculosis useful in diagnostic skin
 XX tests and protective or therapeutic vaccines or compositions.
 PS Claim 37; Page 151-152; 299pp; English.
 XX

CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
 CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
 CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
 CC polypeptides fragments, can be used in pharmaceutical compositions or
 CC vaccines to generate a protective or therapeutic immune response to M.
 CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
 CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
 CC killer cells and/or macrophages in tuberculosis-immune subjects. AA219249
 CC to AA219460 and AA239083 to AA239225 are used in the exemplification of
 CC the present invention

XX SQ Sequence 374 AA;
 Query Match 100.0%; Score 103; DB 2; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AETPGCVAYIGISFLDQASQ 20
 |||||
 Db 239 AETPGCVAYIGISFLDQASQ 258
 |||||

RESULT 14
 AAEL1840
 ID AAEL1840 standard; protein; 374 AA.
 AC AAE11840;
 XX
 XX 18-DEC-2001 (first entry)
 DT
 DE Mycobacterium tuberculosis 38kDa-LP protein.
 XX
 KW 38kDa-LP protein; antibiotic; vaccine; CD8 T-cell; epitope; medicament;
 KW prophylactic; therapeutic; mycobacterial infection.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO200170764-A2.
 XX
 XX 27-SEP-2001.
 PD
 XX
 XX 20-MAR-2001; 2001WO-GB001205.
 PF
 XX
 XX 20-MAR-2000; 2000GB-00006692.
 PR
 XX (GLAX) GLAXO GROUP LTD.
 FA
 XX Klein MR;
 XX
 XX WPI; 2001-607517/69.
 DR
 DR N-PSDB; AAD18855.
 XX

PT Use of polypeptide comprising CD8 T-cell epitopes derived from
 PT Mycobacterium tuberculosis proteins in the manufacture of a medicament
 PT for vaccinating prophylactically or therapeutically against mycobacterial
 PT infection.
 XX
 XX Claim 3; Page 50-52; 54pp; English.
 PS
 XX

CC The invention relates to use of polypeptides comprising CD8 T-cell
 CC epitopes derived from Mycobacterium tuberculosis proteins. The invention
 CC is useful in the manufacture of a medicament for prophylactic or
 CC therapeutic vaccination against mycobacterial infection. The polypeptide
 CC and the expression vector are capable of stimulating CD8 T-cell response.
 CC The invention also provides a vaccine composition which comprises
 CC polypeptide or expression vector useful for vaccinating a pre-selected
 CC host against a mycobacterial infection. The invention further relates to

CC a method for determining the presence or absence of CD8 T-cell response
 CC to epitope sequence by identifying the presence or absence of T-cell that
 CC recognise the epitope sequence in a sample from the host. The present
 CC sequence is Mycobacterium tuberculosis 38kDa-LP protein related to the
 CC invention

XX SQ Sequence 374 AA;
 Query Match 100.0%; Score 103; DB 4; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AETPGCVAYIGISFLDQASQ 20
 |||||
 Db 239 AETPGCVAYIGISFLDQASQ 258
 |||||

RESULT 15
 AAEL29719
 ID AAEL29719 standard; protein; 374 AA.
 XX
 AC AAE29719;
 XX
 XX 27-JAN-2003 (first entry)
 DT
 DE Mycobacterium sp. 38kD antigenic protein.
 XX
 KW Vaccine; immunity; diagnostic agent; gene therapy; 38kD antigen.
 XX
 OS Mycobacterium sp.
 XX

Key Location/Qualifiers
 FT Active-site 1..2
 FT Binding-site 1..2 /label= d
 FT Modified-site 1..2
 FT Cleavage-site 1..2
 FT Cross-links 1..2
 FT Disulfide-bond 1..2
 FT Domain 1..2
 FT Misc-difference 1..2
 FT note= "Encoded by "
 XX
 PN WO200272792-A2.
 XX
 XX 19-SEP-2002.
 PD
 XX
 XX 13-MAR-2002; 2002WO-US008223.
 PF
 XX
 XX 13-MAR-2001; 2001US-0275837P.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX Skeiky Y, Brannon M, Guderian J;
 PI
 XX WPI; 2002-759844/82.
 DR
 DR N-PSDB; AAD47096.
 XX

PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.
 XX
 XX Disclosure; Page 109-110; 155pp; English.
 PS
 XX

CC The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic

CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides, as
CC in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is
CC Mycobacterium sp. 38kD antigenic protein
XX

SQ Sequence 374 AA;

Query Match 100.0%; Score 103; DB 5; Length 374;
Best Local Similarity 100.0%; Pred. No. 3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
|||
Db 239 AETPGCVAYIGISFLDQASQ 258
|||

Search completed: March 10, 2004, 12:05:31
Job time : 48.5641 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 12:11:07 ; Search time 25.1282 Seconds
(without alignments)
168.061 Million cell updates/sec

Title: US-10-044-703-52

Perfect score: 103

Sequence: 1 AETPGCVAYIGISFLDQASQ 20

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103	100.0	20	9	US-09-813-333-52
2	103	100.0	20	13	US-10-044-703-52
3	103	100.0	374	9	US-09-287-849-6
4	103	100.0	374	9	US-09-287-849-40
5	103	100.0	374	14	US-10-193-002-148
6	103	100.0	374	14	US-10-193-002-150
7	103	100.0	374	14	US-10-084-843-153
8	103	100.0	374	14	US-10-084-843-155
9	103	100.0	374	14	US-10-359-460-6
10	103	100.0	374	14	US-10-359-460-40
11	103	100.0	374	14	US-10-098-732A-39
12	103	100.0	652	14	US-10-193-002-350
13	103	100.0	652	14	US-10-084-843-355
14	103	100.0	802	9	US-09-287-849-10
15	103	100.0	802	14	US-10-193-002-209

Sequence 146, App
Sequence 214, App
Sequence 351, App
Sequence 10, Appl
Sequence 11724, A
Sequence 445, App
Sequence 2814, App
Sequence 60, Appl
Sequence 8, Appl
Sequence 1, Appl
Sequence 145, App
Sequence 31, Appl
Sequence 593, App
Sequence 271, App
Sequence 7, Appl
Sequence 5, Appl
Sequence 2, Appl
Sequence 472, Appl
Sequence 10519, A
Sequence 22631, A
Sequence 405, App
Sequence 136, App
Sequence 13667, A
Sequence 720, App
Sequence 3, Appl
Sequence 14063, A
Sequence 2418, App
Sequence 237, App
Sequence 149, App

ALIGNMENTS

RESULT 1

US-09-813-333-52
; Sequence 52, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 52
; TYPE: PRT
; LENGTH: 20
; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-52

Query Match 100.0%; Score 103; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AETPGCVAYIGISFLDQASQ 20

Db 1 AETPGCVAYIGISFLDQASQ 20

RESULT 2

US-10-044-703-52
; Sequence 52, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703

;/ CURRENT FILING DATE: 2002-05-20
;/ PRIOR APPLICATION NUMBER: 60/190,834
;/ PRIOR FILING DATE: 2000-03-20
;/ NUMBER OF SEQ ID NOS: 81
;/ SOFTWARE: PatentIn Ver. 2.1
;/ SEQ ID NO 52
;/ LENGTH: 20
;/ TYPE: PRT
;/ ORGANISM: Mycobacterium tuberculosis
US-10-044-703-52

Query Match 100.0%; Score 103; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
| | | | | | | | | | | | | | | | | | | | | |
DB 1 AETPGCVAYIGISFLDQASQ 20

RESULT 3
US-09-287-849-6
;/ Sequence 6, Application US/09287849
;/ Patent No. US20020009459A1
;/ GENERAL INFORMATION:
;/ APPLICANT: Reed, Steven G.
;/ APPLICANT: Skeiky, Yasir A.W.
;/ APPLICANT: Dillon, Davin C.
;/ APPLICANT: Alderson, Mark
;/ APPLICANT: Campos-Neto, Antonio
;/ APPLICANT: Corixa Corporation
;/ TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
;/ FILE REFERENCE: 014058-009020US
;/ CURRENT APPLICATION NUMBER: US/09/287,849
;/ CURRENT FILING DATE: 1999-04-07
;/ PRIOR APPLICATION NUMBER: US 08/818,112
;/ PRIOR FILING DATE: 1997-03-13
;/ PRIOR APPLICATION NUMBER: US 08/942,578
;/ PRIOR FILING DATE: 1997-10-01
;/ PRIOR APPLICATION NUMBER: US 09/025,197
;/ PRIOR FILING DATE: 1998-02-18
;/ PRIOR APPLICATION NUMBER: US 09/056,556
;/ PRIOR FILING DATE: 1998-04-07
;/ PRIOR APPLICATION NUMBER: US 09/223,040
;/ PRIOR FILING DATE: 1998-12-30
;/ NUMBER OF SEQ ID NOS: 46
;/ SOFTWARE: PatentIn Ver. 2.1
;/ SEQ ID NO 6
;/ LENGTH: 374
;/ TYPE: PRT
;/ ORGANISM: Artificial Sequence
;/ FEATURE:
;/ OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-6

Query Match 100.0%; Score 103; DB 9; Length 374;
Best Local Similarity 100.0%; Pred. No. 4.5e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
| | | | | | | | | | | | | | | | | | | | | |
DB 239 AETPGCVAYIGISFLDQASQ 258

RESULT 4
US-09-287-849-40
;/ Sequence 40, Application US/09287849
;/ Patent No. US20020009459A1
;/ GENERAL INFORMATION:
;/ APPLICANT: Reed, Steven G.
;/ APPLICANT: Skeiky, Yasir A.W.
;/ APPLICANT: Dillon, Davin C.

;/ APPLICANT: Alderson, Mark
;/ APPLICANT: Campos-Neto, Antonio
;/ APPLICANT: Corixa Corporation
;/ TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
;/ FILE REFERENCE: 014058-009020US
;/ CURRENT APPLICATION NUMBER: US/09/287,849
;/ CURRENT FILING DATE: 1999-04-07
;/ PRIOR APPLICATION NUMBER: US 08/818,112
;/ PRIOR FILING DATE: 1997-03-13
;/ PRIOR APPLICATION NUMBER: US 08/942,578
;/ PRIOR FILING DATE: 1997-10-01
;/ PRIOR APPLICATION NUMBER: US 09/025,197
;/ PRIOR FILING DATE: 1998-02-18
;/ PRIOR APPLICATION NUMBER: US 09/056,556
;/ PRIOR FILING DATE: 1998-04-07
;/ PRIOR APPLICATION NUMBER: US 09/223,040
;/ PRIOR FILING DATE: 1998-12-30
;/ NUMBER OF SEQ ID NOS: 46
;/ SOFTWARE: PatentIn Ver. 2.1
;/ SEQ ID NO 40
;/ LENGTH: 374
;/ TYPE: PRT
;/ ORGANISM: Mycobacterium tuberculosis
;/ FEATURE:
;/ OTHER INFORMATION: 38 kD antigen
US-09-287-849-40

Query Match 100.0%; Score 103; DB 9; Length 374;
Best Local Similarity 100.0%; Pred. No. 4.5e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
| | | | | | | | | | | | | | | | | | | | | |
DB 239 AETPGCVAYIGISFLDQASQ 258

RESULT 5
US-10-193-002-148
;/ Sequence 148, Application US/10193002
;/ Publication No. US20030135026A1
;/ GENERAL INFORMATION:
;/ APPLICANT: Reed, Steven G.
;/ Skeiky, Yasir A.W.
;/ Dillon, Davin C.
;/ Campos-Neto, Antonio
;/ Houghton, Raymond
;/ Vedvick, Thomas S.
;/ Twardzik, Daniel R.
;/ Lodes, Michael J.
;/ Hendrickson, Ronald C.
;/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
;/ TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSER: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/10/193,002
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998

```
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 148:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 148:
US-10-193-002-148
Query Match 100.0%; Score 103; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 4.5e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AETPGCVAYIGISFLDQASQ 20
Db 239 AETPGCVAYIGISFLDQASQ 258

RESULT 6
US-10-193-002-150
Sequence 150, Application US/10193002
Publication No. US20030135026A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, David C.
Campos-Neto, Antonia
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 150:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
```

```
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 150:
US-10-193-002-150
Query Match 100.0%; Score 103; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 4.5e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AETPGCVAYIGISFLDQASQ 20
Db 239 AETPGCVAYIGISFLDQASQ 258

RESULT 7
US-10-084-843-153
Sequence 153, Application US/10084843
Publication No. US20030143243A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, David C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 153:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 153:
US-10-084-843-153
Query Match 100.0%; Score 103; DB 14; Length 374;
Best Local Similarity 100.0%; Pred. No. 4.5e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AETPGCVAYIGISFLDQASQ 20
Db 239 AETPGCVAYIGISFLDQASQ 258
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```


REFERENCE/DOCKET NUMBER: 210121.411C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 355:

SEQUENCE CHARACTERISTICS:

LENGTH: 652 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 355:

US-10-084-843-355

Query Match 100.0%; Score 103; DB 14; Length 652;

Best Local Similarity 100.0%; Pred. No. 8.3e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AETPGCVAYIGISFLDQASQ 20

Db 223 AETPGCVAYIGISFLDQASQ 242

RESULT 14

US-09-287-849-10

Sequence 10, Application US/09287849

Patent No. US2002009459A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

APPLICANT: Alderson, Mark

APPLICANT: Campos-Neto, Antonio

APPLICANT: Corixa Corporation

TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens

FILE OF INVENTION: and Their Uses

FILE REFERENCE: 014058-009020US

CURRENT APPLICATION NUMBER: US/09/287,849

CURRENT FILING DATE: 1999-04-07

PRIOR APPLICATION NUMBER: US 08/818,112

PRIOR FILING DATE: 1997-03-13

PRIOR APPLICATION NUMBER: US 08/942,578

PRIOR FILING DATE: 1997-10-01

PRIOR APPLICATION NUMBER: US 09/025,197

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 09/056,556

PRIOR FILING DATE: 1998-04-07

PRIOR APPLICATION NUMBER: US 09/223,040

PRIOR FILING DATE: 1998-12-30

NUMBER OF SEQ ID NOS: 46

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 10

LENGTH: 802

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion

US-09-287-849-10

Query Match

Best Local Similarity 100.0%; Score 103; DB 9; Length 802;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AETPGCVAYIGISFLDQASQ 20

Db 289 AETPGCVAYIGISFLDQASQ 308

RESULT 15

US-10-193-002-209

Sequence 209, Application US/10193002

Publication No. US20030135026A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Wedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS

NUMBER OF SEQUENCES: 350

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/193,002

FILING DATE: 10-Jul-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,596

FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.417C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 209:

SEQUENCE CHARACTERISTICS:

LENGTH: 802 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 209:

US-10-193-002-209

Query Match 100.0%; Score 103; DB 14; Length 802;

Best Local Similarity 100.0%; Pred. No. 1e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AETPGCVAYIGISFLDQASQ 20

Db 289 AETPGCVAYIGISFLDQASQ 308

Search completed: March 10, 2004, 12:41:31

Job time : 25.1282 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:58:01 ; Search time 10.7692 Seconds
(without alignments)
178.641 Million cell updates/sec

Title: US-10-044-703-52

Perfect score: 103

Sequence: 1 AETPGCVAYIGISFLDQASQ 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.*

1: Piri:*

2: Piri:*

3: Piri:*

4: Piri:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	100.0	374	1 F70584	phosphate specific
2	52	50.5	302	2 AC1387	Phosphate ABC tran
3	52	50.5	302	2 AE1762	Phosphate ABC tran
4	50	48.5	90	2 T41548	atpase inhibitor p
5	50	48.5	298	2 A86840	hypothetical prote
6	47	45.6	300	1 A69956	phosphate ABC tran
7	45	43.7	462	2 TC1732	UTP-glucose glucos
8	45	43.7	480	2 B85014	probable flavonol
9	44.5	43.2	173	2 A95864	hypothetical prote
10	44	42.7	192	2 H86543	hypothetical prote
11	44	42.7	192	2 D72081	conserved hypothet
12	44	42.7	292	2 A95163	hypothetical prote
13	44	42.7	292	2 H98028	hypothetical prote
14	43	41.7	38	2 B82858	hypothetical prote
15	43	41.7	259	2 AG3607	probable hydroxlas
16	42	40.8	282	2 A86976	crocidine 5'-phosp
17	42	40.8	283	2 H86839	hypothetical prote
18	42	40.8	305	2 T24534	hypothetical prote
19	42	40.8	325	2 T25331	hypothetical prote
20	42	40.8	365	2 T25490	hypothetical prote
21	42	40.8	369	2 G95100	membrane protein (
22	42	40.8	369	2 A97969	conserved hypothet
23	42	40.8	370	1 W2M139	E2 protein human
24	42	40.8	387	2 T32139	hypothetical prote
25	42	40.8	429	2 C87183	phosphate-binding
26	42	40.8	429	2 T28950	hypothetical prote
27	42	40.8	469	2 D86144	protein probable U
28	42	40.8	656	1 S59631	endo-1,4-beta-xyla
29	42	40.8	1723	2 S58880	receptor DSC-205 -

30 42 40.8 2284 1 GNVVGV genome polypoteine
31 41 39.8 298 2 AE3234 transcripional re
32 41 39.8 312 2 B72056 hypothetical prote
33 41 39.8 312 2 E86559 hypothetical prote
34 41 39.8 352 2 T25170 hypothetical prote
35 41 39.8 371 2 T42623 probable sugar tra
36 41 39.8 546 2 T40888 probable glucose t
37 41 39.8 1146 2 S59376 hypothetical prote
38 40.5 39.3 1383 2 T13052 guanine nucleotide
39 40 38.8 100 2 G72729 hypothetical prote
40 40 38.8 263 2 AB0934 probable membrane
41 40 38.8 264 2 T31594 hypothetical prote
42 40 38.8 280 2 AH1305 metal cations ABC
43 40 38.8 280 2 AH1677 similar metal cati
44 40 38.8 288 2 B91063 hypothetical prote
45 40 38.8 288 2 F85907 hypothetical prote

ALIGNMENTS

RESULT 1

F70584
phosphate specific transporter S precursor - Mycobacterium tuberculosis (strain H37RV)
C:Alternate names: antigen b
C:Species: Mycobacterium tuberculosis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: F70584; J05103; A42930; A49721; A45820
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70584
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-374 <COL>
A:Cross-references: GB:Z95209; GB:AL123456; NID:G3261750; PIDN:CA808484.1; PID:G2078049
R: Braibant, M.; Lefevre, P.; de Wit, L.; Peirs, P.; Ooms, J.; Huygen, K.; Andersen, A.B.; Gene 176, 171-176, 1996
A:Title: A Mycobacterium tuberculosis gene cluster encoding proteins of a phosphate trans
A:Reference number: JC5100; MUID:97075926; PMID:8918249
A:Accession: JC5103
A:Molecule type: DNA
A:Residues: 1-374 <BRA>
A:Cross-references: GB:M30046; NID:G149987; PIDN:AAA25374.1; PID:G149988
A:Note: neither the complete nucleic acid sequence nor the complete translation are show
A:Note: the authors translated the initiation codon GTG for residue 1 as Met
R: Andersen, A.B.; Hansen, E.B.
Infect. Immun. 57, 2481-2486, 1989
A:Title: Structure and mapping of antigenic domains of protein antigen b, a 38,000-molecu
A:Reference number: A42930; MUID:89307568; PMID:2545626
A:Accession: A42930
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-374 <AND1>
A:Cross-references: GB:M30046; NID:G149987; PIDN:AAA25374.1; PID:G149988
R: Chang, Z.; Choudhary, A.; Lathigra, R.; Quirocho, F.A.
J. Biol. Chem. 269, 1956-1958, 1994
A:Title: The immunodominant 38-kDa lipoprotein antigen of Mycobacterium tuberculosis is
A:Reference number: A49721; MUID:94124544; PMID:8294447
A:Accession: A49721
A:Molecule type: protein
A:Residues: 25-34 <CHA>
R: Andersen, A.B.; Ljungqvist, L.; Olsen, M.
J. Gen. Microbiol. 136, 477-480, 1990
A:Title: Evidence that protein antigen b of Mycobacterium tuberculosis is involved in pho
A:Reference number: A45820; MUID:90362031; PMID:2118164
A:Accession: A45820
A:Molecule type: protein

A:Residues: 'X', 27-28, 'X', 30-34, 'XX', 37 <AND2>
A:Note: Confirmed presence of normal signal and absence of lipoprotein attachment
C:Genetics:
A:Gene: phoS1; ptsS
A:Start codon: GTG
C:Superfamily: phosphate-repressible phosphate-binding protein
C:Keywords: surface antigen
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-37/Product: phosphate specific transporter S #status experimental <MAT>

Query Match 100.0%; Score 103; DB 1; Length 374;
Best Local Similarity 100.0%; Pred. NO. 9e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
|||||:::|
DB 239 AETPGCVAYIGISFLDQASQ 258
|||||:::|

RESULT 2
AC1387
phosphate ABC transporter (binding protein) homolog lmo2499 [imported] - Listeria monocytogenes
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 01-Mar-2002
C:Accession: AC1387
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluerer, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative Genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1387
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-302 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAD00577.1; PID:gl6411987; GSPDB:GN00177
A:Experimental source: strain SGD-e
C:Genetics:
A:Gene: lmo2499
C:Superfamily: Methanobacterium phosphate-binding protein psts

Query Match 50.5%; Score 52; DB 2; Length 302;
Best Local Similarity 43.8%; Pred. NO. 0.42;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLD 16
|||||:::|
DB 204 SETPGAISYLAFLSYD 219
|||||:::|

RESULT 3
AE1762
phosphate ABC transporter (binding protein) homolog lin2642 [imported] - Listeria innocua
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 01-Mar-2002
C:Accession: AE1762
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluerer, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative Genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AE1762
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-302 <GLA>
A:Cross-references: GB:AL592022; PIDN:CA097869.1; PID:gl6415179; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:

A:Gene: lin2642
C:Superfamily: Methanobacterium phosphate-binding protein psts

Query Match 50.5%; Score 52; DB 2; Length 302;
Best Local Similarity 43.8%; Pred. NO. 0.42;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLD 16
|||||:::|
DB 204 SETPGAISYLAFLSYD 219
|||||:::|

RESULT 4
T41548
atpase inhibitor protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jan-2003
C:Accession: T41548
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Seeger, K.; Harris, D.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z22001
A:Accession: T41548
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-90 <WOO>
A:Cross-references: EMBL:AL023794; PIDN:CAA19352.1; GSPDB:GN00068; SPDB:SPCC70.02c
A:Experimental source: strain 972h-; cosmid c70
C:Genetics:
A:Map position: 3
A:Gene: SPDB:SPCC70.02c
C:Superfamily: ATPase inhibitor, mitochondrial

Query Match 48.5%; Score 50; DB 2; Length 90;
Best Local Similarity 50.0%; Pred. NO. 0.27;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 4 POCVAYIGISFLDOAS 19
|||||:::|
DB 9 PACISYRGIRFMSKAS 24
|||||:::|

RESULT 5
A86840
hypothetical protein pstF [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 14-Sep-2001
C:Accession: A86840
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich, Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp.
A:Reference number: A86825; MUID:21235186; PMID:11337471
A:Accession: A86840
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-298 <STO>
A:Cross-references: GB:AB005176; PID:gl2724739; PIDN:AAK05819.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: pstF
C:Superfamily: Methanobacterium phosphate-binding protein psts

Query Match 48.5%; Score 50; DB 2; Length 298;
Best Local Similarity 40.0%; Pred. NO. 0.91;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 2 ETPGCVAYIGISFLD 16
|||||:::|
DB 202 QTPGAISYVALSYD 216
|||||:::|

RESULT 6
A69956
phosphate ABC transporter (binding protein) homolog yggG - Bacillus subtilis

C/Species: *Bacillus subtilis*
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C/Accession: A69956
 R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berteaux, S.; Brouillette, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chao, C.; Cron, S.; Drouin, P.; Egan, K.D.; Errington, J.; Fabis, C.; Ferrari, E.; Ehlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabis, C.; Ferrari, E.; Nature 390, 249-256, 1997
 A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Puma, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Laubert, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seron, T.; Winters, P.; Wipat, A.; Tanaka, T.; Terpstra, P.; Tognoni, K.; Tosato, V.; Uchiyama, K.; Authors: Yoshikawa, H.F.; Zumbach, E.; Yoshikawa, H.; Danchin, A.
 A/Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A/Reference number: A69580; MUID:98044033; PMID:9384377
 A/Accession: A69586
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-300 <KUN>
 A/Cross-references: GB:Z99116; GB:AL009126; NID:92634723; PIDN:CAB14429.1; PID:G2634932
 A/Experimental source: strain 168
 C/Genetics:
 A/Gene: yggG
 C/Superfamily: Methanobacterium phosphate-binding protein pbs
 Query Match 45.6%; Score 47; DB 1; Length 300;
 Best Local Similarity 46.7%; Pred. No. 3;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AETPGCVAYIGISFL 15
 Db 204 ADTPGAIGYAFSFL 218
 |||||:|:|:
 RESULT 7
 T01732
 UTP-glucose glucosyltransferase homolog A IG002N01.15 - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Nov-1999
 C/Accession: T01732
 R/Scheet, P.; Maggi, L.
 submitted to the EMBL Data Library, June 1997
 A/Description: The sequence of A. thaliana IG002N01.
 A/Reference number: Z14407
 A/Accession: T01732
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-462 <SCH>
 A/Cross-references: EMBL:AF007269; NID:G2191126; PID:G2191136
 A/Experimental source: cultivar Columbia
 C/Genetics:
 A/Map position: 4
 A/Introns: 145/2
 A/Note: A IG002N01.15
 C/Superfamily: flavonol O3-glucosyltransferase
 Query Match 43.7%; Score 45; DB 2; Length 462;
 Best Local Similarity 60.0%; Pred. No. 10;
 Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 4 PGCYVAYIGISFLDOA 18
 Db 154 PGCYVAYIGISFLDOA 168
 |||||:|:|:
 RESULT 8
 B85014
 Probable flavonol glucosyltransferase [imported] - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Mar-2001
 C/Accession: B85014
 R/Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Nature 402, 769-777, 1999
 A/Title: Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.
 A/Reference number: A85001; MUID:20083488; PMID:10617198
 A/Accession: B85014
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-480 <STO>
 A/Cross-references: GB:NC_001268; NID:G7267604; PIDN:CAB80916.1; GSPDB:GN00140
 C/Genetics:
 A/Gene: AT4G01070
 A/Map position: 4
 C/Superfamily: flavonol O3-glucosyltransferase
 Query Match 43.7%; Score 45; DB 2; Length 480;
 Best Local Similarity 60.0%; Pred. No. 11;
 Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 4 PGCYVAYIGISFLDOA 18
 Db 172 PGCYVAYIGISFLDOA 186
 |||||:|:|:
 RESULT 9
 A95864
 hypothetical protein [imported] - Sinorhizobium meliloti (strain 1021) megaplasmid pSymB
 C/Species: Sinorhizobium meliloti
 C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C/Accession: A95864
 R/Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan, Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A/Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo:
 A/Reference number: A95842; MUID:21396508; PMID:11481431
 A/Accession: A95864
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-173 <KUR>
 A/Cross-references: GB:AL591985; PIDN:CAC48577.1; PID:G15140049; GSPDB:GN00167
 A/Experimental source: strain 1021, megaplasmid pSymB
 R/Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, habault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.C
 A/Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A/Reference number: A96039; MUID:21368234; PMID:11474104
 A/Contents: annotation
 C/Genetics:
 A/Gene: SMB20177
 A/Genome: plasmid
 Query Match 43.2%; Score 44.5; DB 2; Length 173;
 Best Local Similarity 42.1%; Pred. No. 4.6;
 Matches 8; Conservative 8; Mismatches 2; Indels 1; Gaps 1;
 QY 3 TPGC-VAYIGISFLDOA 20
 Db 30 TPGC-VAYIGISFLDOA 48
 |||||:|:|:
 RESULT 10
 H86543
 hypothetical protein CPJ0426 [imported] - Chlamydomonas reinhardtii (strain J138)
 C/Species: Chlamydomonas reinhardtii, Chlamydomonas reinhardtii
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
 C/Accession: H86543
 R/Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Iet Nucleic Acids Res 28, 2311-2314, 2000
 A/Title: Comparison of whole genome sequences of *Chlamydomonas reinhardtii* J138.
 A/Reference number: A86491; MUID:20330349; PMID:10871362

A:Accession: H86543
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <STO>
A:Cross-references: GB:BA000008; NID:g8978798; PIDN:BAA98634.1; GSPDB:GN00142
C:Genetics:
A:Gene: CPJ0426

Query Match 42.7%; Score 44; DB 2; Length 192;
Best Local Similarity 77.8%; Pred. No. 6.2;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ETPGCYAYI 10
||| |||
Db 106 ETPSCIAYI 114

RESULT 11
D72081
conserved hypothetical protein frameshifted CP0327 [imported] - Chlamydomonadales pneumoniae
N:Alternate names: hypothetical protein CT277 homolog
C:Species: Chlamydomonadales pneumoniae, Chlamydomonadales pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: D72081; G81589
R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative Genomes of Chlamydomonadales pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: D72081
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <ARN>
A:Cross-references: GB:AE001625; GB:AE001363; NID:g4376695; PIDN:AA18570.1; PID:g4376700
A:Experimental source: strain CWL029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwynn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydomonadales pneumoniae and Chlamydomonadales pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: G81589
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <REA>
A:Cross-references: GB:AE002195; GB:AE002161; NID:g7189246; PIDN:AAF38182.1; PID:g7189252
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CPN0426; CP0327

Query Match 42.7%; Score 44; DB 2; Length 192;
Best Local Similarity 77.8%; Pred. No. 6.2;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ETPGCYAYI 10
||| |||
Db 106 ETPSCIAYI 114

RESULT 12
A95163
hypothetical protein SPI400 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: A95163
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Hickey, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
Science 293, 498-506, 2001
A:Authors: Lofcus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: A95163
A:Status: preliminary
A:Molecule type: DNA

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A.; Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tshahko, M.H.; Wallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A.; Reference number: A59328
 A;Contents: annotation
 C;Genetics:
 A;Gene: XFO014

Query Match 41.7%; Score 43; DB 2; Length 38;
 Best Local Similarity 53.3%; Pred. No. 1.7;
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Cy 4 PGCYVYIGISFLDQA 18
 ||: :||: |||
 Db 22 PGVMSGMGITFLDSA 36

RESULT 15

AG3607
 Probable hydroxylase [imported] - Brucella melitensis (strain 16M)
 C;Species: Brucella melitensis
 C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 C;Accession: AG3607
 R;DeiVecchio, V.G.; Kaputal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
 .; Mazur, M.; Goltzman, E.; Salkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
 A;Reference number: AD3252; PMID:11756688
 A;Accession: AG3607
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-259 <KUR>
 A;Cross-references: GB:AE008918; PIDN:AAL54026.1; PID:gi7984978; GSPDB:GN00191
 A;Experimental source: strain 16M
 C;Genetics:
 A;Gene: BMEI10784
 A;Map position: II

Query Match 41.7%; Score 43; DB 2; Length 259;
 Best Local Similarity 41.2%; Pred. No. 13;
 Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Cy 4 PGCYVYIGISFLDQASQ 20
 ||: :||: :||: ||:
 Db 72 PGVLGYIGVADVDAAAE 88

Search completed: March 10, 2004, 12:12:28
 Job time : 12.7692 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:51:36 ; Search time 6.15385 Seconds
(without alignments)
169.228 Million cell updates/sec

Title: US-10-044-703-52

Perfect score: 103

Sequence: 1 AETPGCVAYIGISFLDQASQ 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103	100.0	374	1	PST1_MYCTU
2	50	48.5	90	1	IATP_SCHPO
3	47	45.6	300	1	YQGG_BACSU
4	45	43.7	480	1	HQGT_ARATH
5	43	41.7	191	1	PGHD_FELCA
6	43	41.7	470	1	HQGT_RAUSE
7	42	40.8	282	1	PYRF_MYCLE
8	42	40.8	370	1	VE2_HPV39
9	42	40.8	1722	1	LY75_HUMAN
10	42	40.8	1722	1	LY75_MESAU
11	42	40.8	1723	1	LY75_MOUSE
12	42	40.8	2284	1	POL1_GFLV
13	41.5	40.3	942	1	MSH2_MAIZE
14	41	39.8	546	1	GHTS_SCHPO
15	40.5	39.3	1571	1	C3G_DROME
16	40	38.8	263	1	YPUD_ECOLI
17	40	38.8	280	1	MNTC_LISIN
18	40	38.8	280	1	MNTC_LISMO
19	40	38.8	362	1	MURG_STRCU
20	40	38.8	364	1	MURG_STRCO
21	40	38.8	356	1	ASSY_STRAS
22	40	38.8	396	1	ASSY_STRAS
23	40	38.8	426	1	Y680_CHLPN
24	40	38.8	444	1	ADRL_DROME
25	40	38.8	874	1	GLND_VIBPA
26	40	38.8	931	1	SM6C_MOUSE
27	40	38.8	960	1	SM6C_RAT
28	40	38.8	1277	1	NPC1_PIG
29	39.5	38.3	324	1	OSA_DROSI
30	39.5	38.3	324	1	OSA_DROYA
31	39.5	38.3	2716	1	OSA_DROME
32	39	37.9	357	1	YAUH_SCHPO
33	39	37.9	513	1	TRPE_BACPU

RESULT 1

PST1_MYCTU
AC P15712; O05868; STANDARD; PRT; 374 AA.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phosphate-binding protein 1 precursor (PBP-1) (PstS-1) (Protein antigen B) (PAB) (Antigen Ag78).
GN PSTS1 OR PHOS1 OR RV0934 OR MT0961 OR MTCY08D3.05C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=89307568; PubMed=2545626;
RX Andersen A.B., Hansen E.B.;
RT "Structure and mapping of antigenic domains of protein antigen b, a 38,000-molecular-weight protein of Mycobacterium tuberculosis.";
RL Infect. Immun. 57:2481-2488(1989).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaija F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence";
RL Nature 393:537-544(1998).
RN [3]
SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E., Kolony J.F., Nelson W.C., Umayam L.A., Emdalaeva M., Salzberg S.L., Bishai W., Jacobs W.R. Jr., Weidman J., Khouri H., Gill J., Mikula A., "Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- FUNCTION: Required for binding-protein-mediated phosphate transport.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor (Probable).
CC -!- SIMILARITY: Belongs to the pstS family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

34 39 37.9 515 1 TRPE_BACSU
35 39 37.9 537 1 GAG_BAEVM
36 39 37.9 806 1 BIMA_EMENT
37 39 37.9 866 1 BAL_MOUSE
38 39 37.9 873 1 GLND_VIBVU
39 39 37.9 1122 1 DPOL_ADE07
40 39 37.9 1193 1 DPOL_ADE04
41 39 37.9 1835 1 DURI_YEAST
42 38.5 37.4 277 1 HXDB_NORVI
43 38 36.9 227 1 DSG1_MOUSE
44 38 36.9 342 1 CCR6_CERAE
45 38 36.9 342 1 CCR6_HUMAN
P03963 bacillus su
P03341 baboon endo
P17885 emeritella
Q8cas9 mus musculu
Q8dbg3 vibrio vuln
P05664 human adeno
P87503 human adeno
P32528 saccharomyc
P31263 notocpithaim
Q61495 mus musculu
O18983 cercopitheci
O00574 homo sapien

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; M30046; AAA25374.1; -
 DR EMBL; Z95209; CAB08484.1; -
 DR EMBL; AE006981; AAK45208.1; -
 DR PIR; F70584; F70584.
 DR HSP; P06128; 11XH.
 DR TIGR; MT0961; -
 DR TubercuList; Rv0934; -
 DR InterPro; IPR000437; Prok_lipoprot_s.
 DR InterPro; IPR006059; SPP_bac_1.
 DR Pfam; PF01547; SPP_bac_1; 1.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Phosphate transport; transport; Membrane; Lipoprotein; Signal;
 KW Antigen; Complete proteome; Palmitate.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 374 PHOSPHATE-BINDING PROTEIN 1.
 FT LIPID 24 24 N-palmitoyl cysteine (Potential).
 FT LIPID 24 24 S-diacylglycerol cysteine (Potential).
 SQ SEQUENCE 374 AA; 38243 MW; 8334368191FF38A CRC64;

Query Match 100.0%; Score 103; DB 1; Length 374;
 Best Local Similarity 100.0%; Pred. No. 7.1e-10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
 DB 239 AETPGCVAYIGISFLDQASQ 258

RESULT 2

QY IATP SCHPO STANDARD; PRT; 90 AA.
 AC 074523;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative ATPase inhibitor, mitochondrial precursor.
 GN SPC070.02C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RC MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Money P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -!- FUNCTION: Forms a one-to-one complex with ATPase to inhibit the
 CC enzyme activity completely (By similarity).
 CC -!- SUBCELLULAR LOCATION: Mitochondrial (By similarity).
 CC -!- SIMILARITY: Belongs to the ATPase inhibitor family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AL023794; CAA19352.1; -
 DR PIR; T41548; T41548.
 DR GeneDB; SPombe; SPC070.02C; -
 DR InterPro; IPR007648; IATP.
 DR Pfam; PF04568; IATP; 1.
 DR Hypothetical protein; Mitochondrion; Transit peptide.
 KW TRANSIT 1 MITOCHONDRION.
 FT CHAIN ? 90 PUTATIVE ATPASE INHIBITOR.
 SQ SEQUENCE 90 AA; 10674 MW; F4C4DF852E3E6909 CRC64;

Query Match 48.5%; Score 50; DB 1; Length 90;
 Best Local Similarity 50.0%; Pred. No. 0.17;
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 4 PGCVAYIGISFLDQAS 19
 DB 9 PACISYRGIRFMSKAS 24

RESULT 3

QY YGGG BACSU
 ID YQGG BACSU STANDARD; PRT; 300 AA.
 AC P46338;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable ABC transporter binding protein YQGG precursor.
 GN YQGG OR BSU24990.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / JH642;
 RC MEDLINE=96349106; PubMed=8760913;
 RA Takemaru K.-I., Mizuno M., Kobayashi Y.;
 RT "A Bacillus subtilis gene cluster similar to the Escherichia coli
 RT phosphate-specific transport (pst) operon: evidence for a tandemly
 RT arranged pstB gene.";
 RL Microbiology 142:2017-2020(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / JH642;
 RC MEDLINE=97124195; PubMed=8969508;
 RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
 RA Kobayashi Y.;
 RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
 RT the Bacillus subtilis genome containing the skin element and many
 RT sporulation genes.";
 RL Microbiology 142:3103-3111(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RC MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

28-FEB-2003 (Rel. 41, Last annotation update)
 DN Hydroquinone glucosyltransferase (EC 2.4.1.218) (Arbutin synthase).
 AS
 OS Rauvolfia serpentina (Serpentwood) (Devilpepper).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 CC lamids; Gentianales; Apocynaceae; Rauvolfioideae; Vinaceae;
 CC Rauvolfia.
 CC NCBI_TaxID=4060;
 RN [1]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RA MEDLINE=21934706; PubMed=1937332;
 RX Heffner T., Arend J., Warzecha H., Siems K., Stoeckigt J.;
 RT "Arbutin synthase, a novel member of the NR1beta glucosyltransferase
 RT family, is a unique multifunctional enzyme converting various natural
 RT products and xenobiotics.";
 RL Bioorg. Med. Chem. 10:1731-1741(2002).
 RN [2]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RA MEDLINE=2014422; PubMed=10680170;
 RX Arend J., Warzecha H., Stoeckigt J.;
 RT "Hydroquinone-O-glucosyltransferase from cultivated Rauvolfia cells:
 RT enrichment and partial amino acid sequences.";
 RL Phytochemistry 53:187-193(2000).
 CC -1- FUNCTION: Broad spectrum multifunctional glucosyltransferase. In
 CC addition to hydroquinone it accept at least 45 natural and
 CC synthetic phenols as well as two cinnamyl alcohols as substrates.
 CC Hydroquinone was however the best substrate. In contrast to this
 CC broad acceptor substrate specificity, only pyrimidine nucleotide
 CC activated glucose is tolerated as a donor substrate.
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + hydroquinone = UDP +
 CC hydroquinone-O-beta-D-glucopyranoside.
 CC -1- SIMILARITY: Belongs to the UDP-glucosyltransferase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ310148; CAC35167.1;
 DR InterPro: IPR002213; UDP_gluco_trans.
 DR Pfam; PF00201; UDPGT, 1.
 DR PROSITE; PS00375; UDPGT, 1.
 KW Transferase; Glycosyltransferase.
 SQ SEQUENCE 470 AA; 51793 MW; 9C1B4A0760321F51 CRC64;
 Query Match 41.7%; Score 43; DB 1; Length 470;
 Best Local Similarity 47.1%; Pred. No. 13;
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 2 ETGCVAYIGISTLQQA 18
 Db 167 QIPGCIPIHGKDFLQPA 183
 RESULT 7
 ID PYRF MYCLE STANDARD; PRT; 282 AA.
 AC QSCCEI;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMP decarboxylase)
 DE (OMPDCase) (OMPDecase).
 DE PYRF OR ML0537.
 GN Mycobacterium leprae.
 OS Mycobacterium leprae.
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 CC NCBI_TaxID=1769;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN:TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Biglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus.";
 RL Nature 409:1007-1011(2001).
 CC -1- CATALYTIC ACTIVITY: Orotidine 5'-phosphate = UMP + CO(2).
 CC -1- PATHWAY: Pyrimidine biosynthesis; sixth (last) step.
 CC -1- SIMILARITY: Belongs to the OMP decarboxylase family. Subfamily 2.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL583918; CAC30045.1;
 DR PIR; A86976; A86976.
 DR Leproma; ML0537;
 DR HAMAP; MF_01215; -;
 DR InterPro: IPR001754; OMPDecase.
 DR Pfam; PF00215; OMPDecase; 1.
 DR PROSITE; PS00156; OMPDecase; 1.
 KW Lyase; Decarboxylase; Pyrimidine biosynthesis; Complete proteome.
 FT ACT SITE 95 PROTON DONOR (BY SIMILARITY).
 SQ SEQUENCE 282 AA; 28678 MW; 89DE23C44B27BA5 CRC64;
 Query Match 40.8%; Score 42; DB 1; Length 282;
 Best Local Similarity 55.6%; Pred. No. 12;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 4 PGCVAYIGI 12
 Db 194 PGCLGYGVG 202
 RESULT 8
 ID VE2 HPV39 STANDARD; PRT; 370 AA.
 AC P24830;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Regulatory protein E2.
 GN E2.
 OS Human papillomavirus type 39.
 CC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 CC Papillomavirus.
 CC NCBI_TaxID=10588;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91135017; PubMed=1847266;
 RA Volpers C., Strecek R.B.;
 RT "Genome organization and nucleotide sequence of human papillomavirus
 RT type 39.";
 RL Virology 181:419-423(1991).
 CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACNNNNNGGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA

```
CC REPLICATION.
```

```
CC -!- SUBUNIT: Binds DNA as a dimer.
```

```
CC -!- SUBCELLULAR LOCATION: Nuclear.
```

```
-----
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@isb-sib.ch).
```

```
-----
```

```
DR EMBL; M62849; AAA47053.1; .
```

```
DR PIR; D38502; W2WL39.
```

```
DR HSRF; PL7383; LDHM.
```

```
DR InterPro; IPR000427; E2_C.
```

```
DR InterPro; IPR001866; E2_N.
```

```
DR InterPro; IPR003021; Viral_DNA_bd.
```

```
DR Pfam; PF00511; E2_C; 1.
```

```
DR Pfam; PF00508; E2_N; 1.
```

```
DR ProDom; PD000672; E2_C; 1.
```

```
DR ProDom; PD000678; E2_N; 1.
```

```
DR Early protein; Transcription regulation; Activator; DNA-binding;
```

```
KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
```

```
SQ SEQUENCE 370 AA; 42697 MW; DBEF2CDE1C17F17 CRC64;
```

```
Query Match 40.8%; Score 42; DB 1; Length 370;
```

```
Best Local Similarity 43.8%; Pred.No. 16;
```

```
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 ETGCGVAYIGISFLDQ 17  
: ||||| :  
Db 153 KTEGCVDWGIYYWNE 168
```

```
RESULT 9
```

```
LY75 HUMAN
```

```
ID LY75 HUMAN STANDARD; PRT; 1722 AA.
```

```
AC O60449; Q75913;
```

```
DT 10-OCT-2003 (Rel. 42, Created)
```

```
DT 10-OCT-2003 (Rel. 42, Last sequence update)
```

```
DT DT 10-OCT-2003 (Rel. 42, Last annotation update)
```

```
DE Lymphocyte antigen 75 precursor (DEC-205) (CD205 antigen) (gp200-MR6).
```

```
GN LY75 OR CD205.
```

```
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI_TaxID=9606;  
RN [1]
```

```
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, TISSUE SPECIFICITY, AND  
RP GLYCOSYLATION.
```

```
RC TISSUE=Thymus;
```

```
RX MEDLINE=99077159; PubMed=9862343;
```

```
RA McKay P.F., Imami N., Johns M., Taylor-Fishwick D.A., Sedibane L.M.,  
RA Torty N.F., Hsuan J.J., Palmer D.B., George A.J.T., Foxwell B.M.J.,  
RA Ritter M.A.;
```

```
RT "The gp200-Mr6 molecule which is functionally associated with the IL-4  
RT receptor modulates B cell phenotype and is a novel member of the human  
RT macrophage mannose receptor family.";  
RL Eur. J. Immunol. 28:4071-4083(1998).  
[2]
```

```
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=98221117; PubMed=9553150;
```

```
RA Kato M., Neil T.K., Clark G.J., Morris C.M., Sorg R.V., Hart D.N.J.;  
RT "cDNA cloning of human DEC-205, a putative antigen-uptake receptor on  
RT dendritic cells.";  
RL Immuno genetics 47:442-450(1998).
```

```
CC -!- FUNCTION: Acts as an endocytic receptor to direct captured  
CC antigens from the extracellular space to a specialized antigen-  
CC processing compartment (By similarity). Causes reduced  
CC proliferation of B lymphocytes.
```

```
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
```

```
CC -!- TISSUE SPECIFICITY: Expressed in spleen, thymus, colon and
```

FT CONFLICT 807 807 D -> E (IN REF. 1).
 FT CONFLICT 1321 1321 N -> K (IN REF. 1).
 FT CONFLICT 1391 1391 H -> Y (IN REF. 1).
 SQ CONFLICT 1722 AA; 19820 MW; 3F9D77759C2FE4A5 CRC64;
 Query Match 40.8%; Score 42; DB 1; Length 1722;
 Best Local Similarity 60.0%; Pred. No. 71;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 2 ETPGCVAYIG 11
 Db 458 KTPNCVSYLG 467

RESULT 10
 LY75 MESAU STANDARD; PRT; 1722 AA.
 AC Q20P9;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Lymphocyte antigen 75 precursor (DEC-205).
 GN LY75.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
 RC TISSUE=Bone marrow;
 RX MEDLINE=22165779; PubMed=12175539;
 RA Maruyama K., Akiyama Y., Cheng J., Nara-Ashizawa N., Hojo T.,
 RA Sasaki K., Yamaguchi K.;
 RT "Hamster DEC-205, its primary structure, tissue and cellular
 distribution."
 RL Cancer Lett. 181:223-232(2002).
 CC -1- FUNCTION: Acts as an endocytic receptor to direct captured
 antigens from the extracellular space to a specialized antigen-
 processing compartment. Causes reduced proliferation of B
 lymphocytes (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: Expressed in the thymus and cultured bone
 marrow cells.
 CC -1- SIMILARITY: Contains 9 C-type lectin family domains.
 CC -1- SIMILARITY: Contains 1 fibronectin type II-like domain.
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB059273; BAB69491.1; -
 CC InterPro; IPR000562; FN Type II.
 CC InterPro; IPR001304; LectinC.
 CC InterPro; IPR000772; RicinB_Lectin.
 CC InterPro; IPR008997; RicinB_Like.
 CC Pfam; PF00040; fn2; 1.
 CC Pfam; PF00059; lectin.c; 9.
 CC Pfam; PF00652; RicinB_lectin; 1.
 CC ProDom; PD000995; FN_Type II; 1.
 CC SMART; SM00034; CLECT; 10.
 CC SMART; SM00059; FN2; 1.
 CC SMART; SM00458; RICIN; 1.
 CC PROSITE; PS00615; C-TYPE LECTIN 1; 1.
 CC PROSITE; PS00441; C-TYPE LECTIN 2; 9.
 CC PROSITE; PS00023; FIBRONECTIN 2; FALSE_NEG.
 CC PROSITE; PS02031; RICIN B LECTIN; 1.
 KW Endocytosis; Receptor; Antigen; Repeat; Signal; Transmembrane; Lectin;

KW Glycoprotein. 1 27
 FT SIGNAL 28 1722 BY SIMILARITY.
 FT CHAIN 28 1722 LYMPHOCYTE ANTIGEN 75.
 FT DOMAIN 28 1666 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1667 1691 POTENTIAL.
 FT DOMAIN 1692 1722 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 32 182 RICIN B-TYPE LECTIN.
 FT DOMAIN 169 209 FIBRONECTIN TYPE-II.
 FT DOMAIN 225 341 C-TYPE LECTIN 1.
 FT DOMAIN 368 486 C-TYPE LECTIN 2.
 FT DOMAIN 493 625 C-TYPE LECTIN 3.
 FT DOMAIN 652 791 C-TYPE LECTIN 4.
 FT DOMAIN 958 1091 C-TYPE LECTIN 5.
 FT DOMAIN 1110 1222 C-TYPE LECTIN 6.
 FT DOMAIN 1251 1374 C-TYPE LECTIN 7.
 FT DOMAIN 1401 1513 C-TYPE LECTIN 8.
 FT DOMAIN 1542 1661 C-TYPE LECTIN 9.
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 377 377 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 529 529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 843 843 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 865 865 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 934 934 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1076 1076 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1225 1225 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1320 1320 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1392 1392 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1593 1593 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1626 1626 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1722 AA; 197891 MW; AC80D451927654AF CRC64;
 Query Match 40.8%; Score 42; DB 1; Length 1722;
 Best Local Similarity 60.0%; Pred. No. 71;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 2 ETPGCVAYIG 11
 Db 458 KTPNCVSYLG 467

RESULT 11
 LY75 MOUSE STANDARD; PRT; 1723 AA.
 AC Q60767; Q8C773; Q91XL8; Q9QUZ6;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Lymphocyte antigen 75 precursor (DEC-205) (CD205 antigen).
 GN LY75 OR CD205.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, TISSUE SPECIFICITY, AND
 GLYCOSYLATION.
 RC STRAIN=BALB/c; TISSUE=Dendritic cell, and Thymus;
 RX MEDLINE=95272679; PubMed=7753172;
 RA Jiang W., Swiggard W.J., Heurlier C., Peng M., Mirza A., Steinman R.M.,
 RA Nussenzweig M.C.;
 RT "The receptor DEC-205 expressed by dendritic cells and thymic
 epithelial cells is involved in antigen processing.";
 RL Nature 375:151-155(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Spleen;
 RA Park C.G., Steinman R.M.;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-485 FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=22354683; PubMed=12466851;

RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schorbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gastermond T., Gariboldi M., Glissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierka R.M., King B.A.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.L.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Perle G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Sempie C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tonita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kichikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayaishizaki Y.,
 RA "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 RN [4].
 RP SEQUENCE OF 28-52, SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND
 RP GLYCOSYLATION.
 RC STRAIN=BALB/c; TISSUE=Thymus;
 RX MEDLINE=96021270; PubMed=7553896;
 RA Swiggard W.J., Mirza A., Nussenzweig M.C., Steinman R.M.,
 RT "DSC-205, a 205-kDa protein abundant on mouse dendritic cells and
 RT thymic epithelium that is detected by the monoclonal antibody NLDC-
 RT 145: Purification, characterization, and N-terminal amino acid
 RT sequence";
 RL Cell. Immunol. 165:302-311(1995).
 CC -!- FUNCTION: Acts as an endocytic receptor to direct captured
 CC antigens from the extracellular space to a specialized antigen-
 CC processing compartment. Causes reduced proliferation of B
 CC lymphocytes (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Expressed in dendritic and thymic epithelial
 CC cells and lymph nodes.
 CC -!- PTM: N-glycosylated.
 CC -!- SIMILARITY: Contains 9 C-type lectin family domains.
 CC -!- SIMILARITY: Contains 1 fibronectin type II-like domain.
 CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U19271; AAA80215.1; -;
 CC EMBL; AF395445; AAK81722.1; -;
 CC EMBL; AK049301; BAC33668.1; -;
 CC PIR; S58880; S58880.
 CC HSSP; P02751; 2FN2.
 CC MGD; MGI:106662; Ly75.
 CC InterPro; IPR000562; FN_Type_II.
 CC InterPro; IPR001304; Lectin_C.
 CC InterPro; IPR000772; Ricin_B_lectin.
 CC InterPro; IPR008997; RicinE_like.
 CC Pfam; PF00040; fn2; 1.
 CC Pfam; PF00059; lectin_c; 9.
 CC Pfam; PF00652; Ricin_B_lectin; 1.

DR ProDom; PD000995; FN_Type_II; 1.
 DR SMART; SMO0034; CLECT; 10.
 DR SMART; SMO0059; FN2; 1.
 DR SMART; SMO0458; RICIN; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN 1; 2.
 DR PROSITE; PS00041; C-TYPE LECTIN 2; 9.
 DR PROSITE; PS00023; FIBRONECTIN 2; 1.
 DR PROSITE; PS00231; RICIN B LECTIN; 1.
 KW Endocytosis; Receptor; Antigen; Repeat; Signal; Transmembrane; Lectin;
 KW Glycoprotein.
 FT SIGNAL 1 27
 FT CHAIN 28 1723 LYMPOCYTE ANTIGEN 75.
 FT DOMAIN 28 1667 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1668 1692 POTENTIAL.
 FT DOMAIN 1693 1723 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 33 182 RICIN B-TYPE LECTIN.
 FT DOMAIN 169 209 FIBRONECTIN TYPE-II.
 FT DOMAIN 225 341 C-TYPE LECTIN 1.
 FT DOMAIN 368 486 C-TYPE LECTIN 2.
 FT DOMAIN 493 625 C-TYPE LECTIN 3.
 FT DOMAIN 652 791 C-TYPE LECTIN 4.
 FT DOMAIN 959 1092 C-TYPE LECTIN 5.
 FT DOMAIN 1111 1223 C-TYPE LECTIN 6.
 FT DOMAIN 1252 1375 C-TYPE LECTIN 7.
 FT DOMAIN 1402 1514 C-TYPE LECTIN 8.
 FT DOMAIN 1543 1662 C-TYPE LECTIN 9.
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 377 377 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 843 843 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 865 865 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 935 935 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1077 1077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1104 1104 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1226 1226 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1321 1321 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1393 1393 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1594 1594 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1627 1627 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 51 51 S -> F (IN REF. 4).
 FT CONFLICT 1283 1283 P -> S (IN REF. 2).
 SQ SEQUENCE 1723 AA; 197361 MW; F0621F9420EC04D2 CRC64;
 Query Match 40.8%; Score 42; DB 1; Length 1723;
 Best Local Similarity 60.0%; Pred.No. 71;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 2 BTPGCVAYIG 11
 DB 458 KTPNCVSYLG 467
 RESULT 12
 ID POLI_GFLV STANDARD; PRT; 2284 AA.
 AC P29149;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE RNAL polyprotein (253 kDa protein) [Contains: 63 kDa protease
 DE cofactor; 72 kDa membrane-binding protein; genome-linked protein
 DE (VPg); Protease (EC 3.4.22.-); RNA-directed RNA polymerase
 DE (EC 2.7.7.48)].
 OS Grapevine fanleaf virus (GFLV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
 OC Nepovirus.
 OX NCBI_TaxID=12274;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=F13;
 RC MEDLINE=92013951; PubMed=1655953;
 RA Ritzenthaler C., Viry M., Pinck M., Margis R., Fuchs M., Pinck L.;
 RT "Complete nucleotide sequence and genetic organization of grapevine

```

RT fanleaf nepovirus RNA1.";
RL J. Gen. Virol. 72:2357-2365 (1991).
RN [2]
RX MEDLINE=91285092; PubMed=2060618;
RA Pinck M., Reinbolt J., Ludes A.M., Le Ret M., Pinck L.;
RT "Primary structure and location of the genome-linked protein (VPg) of
RT grapevine fanleaf nepovirus.";
RL FEBS Lett. 284:117-119 (1991).
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D00915; BAA00761.1; -
CC PIR; J01373; GNVVGV.
CC MEROPS; C03.004; -.
CC InterPro; IPR004004; Calici_pol_hel.
CC InterPro; IPR000605; RNA_helicase.
CC InterPro; IPR007095; RNA_pol_DS_Ps.
CC InterPro; IPR001205; RNA_pol_B3D.
CC InterPro; IPR007094; RNA_pol_PSVir.
CC Pfam; PF00680; RNA_dep_RNA_pol; 1.
CC Pfam; PF00910; RNA_helicase; 1.
CC PRINTS; PR00918; CALICIVIRUS.
CC PolyProtein; Transmembrane; Hydrolase; Protease; Transferase;
KW RNA-directed RNA polymerase; ATP-binding.
FT CHAIN 1 568 63 kDa PROTEASE COFACTOR (POTENTIAL).
FT CHAIN 569 1217 72 kDa MEMBRANE-BINDING PROTEIN
FT CHAIN (POTENTIAL).
FT CHAIN 1218 1241 GENOME-LINKED PROTEIN.
FT CHAIN 1242 1460 CYSTEINE PROTEASE (POTENTIAL).
FT CHAIN 1461 2284 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT NP BIND 781 788 ATP (POTENTIAL).
FT ACT SITE 1420 1420 CYSTEINE PROTEASE (POTENTIAL).
SQ SEQUENCE 2284 AA; 252930 MW; 75C6DEB1F45636E9 CRC64;

Query Match 40.8%; Score 42; DB 1; Length 2284;
Best Local Similarity 58.3%; Pred. No. 94;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 5 GCVAYIGSFELD 16 PRT; 942 AA.
Db 168 GAVAYGVNFID 179

RESULT 13
MSH2_MAIZE
ID_MSH2_MAIZE STANDARD; PRT; 942 AA.
AC O9XGC9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNA mismatch repair protein MSH2 (MUS1).
GN MUS1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Horwath M., Kunze R.;
RT "Isolation and characterization of mus1, a putative mismatch repair
RT gene of Zea mays.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

```

```

CC -!- FUNCTION: Involved in postreplication mismatch repair. Binds
CC specifically to DNA containing mismatched nucleotides thus
CC providing a target for the excision repair processes
CC characteristic of postreplication mismatch repair (By similarity).
CC -!- SUBUNIT: Heterodimer of MSH2 and MSH6 (GTBP) (By similarity).
CC -!- SIMILARITY: Belongs to the DNA mismatch repair mutS family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ238785; CAB42554.1; -
CC InterPro; IPR000432; MutS_C.
CC InterPro; IPR007860; MutS_II.
CC InterPro; IPR007696; MutS_III.
CC InterPro; IPR007861; MutS_IV.
CC InterPro; IPR007695; MutS_N.
CC Pfam; PF01624; MutS_I; 1.
CC Pfam; PF05188; MutS_II; 1.
CC Pfam; PF05192; MutS_III; 1.
CC Pfam; PF05190; MutS_IV; 1.
CC Pfam; PF00488; MutS_V; 1.
CC ProDom; PD001263; MutS_C; 1.
CC SMART; SM00534; MUTSac; 1.
CC SMART; SM00533; MUTSd; 1.
CC PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
KW DNA repair; ATP-binding; DNA-binding.
FT NP BIND 667 674 ATP (POTENTIAL).
SQ SEQUENCE 942 AA; 105070 MW; 79B82EB149839C4F CRC64;

Query Match 40.3%; Score 41.5; DB 1; Length 942;
Best Local Similarity 40.0%; Pred. No. 48;
Matches 8; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

Oy 4 PGC---VAYIGSFELDQASQ 20
Db 151 PACRESQVGLSFLDMTNR 170

RESULT 14
GHT5_SCHPO
ID_GHT5_SCHPO STANDARD; PRT; 546 AA.
AC P78831;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE High-affinity glucose transporter ght5 (Hexose transporter 5).
GN GHT5 OR SPCC1235.14.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=972;
RX MEDLINE=2020352; PubMed=10735857;
RA Heiland S., Radovanovic N., Hofer M., Winderickx J., Lichtenberg H.;
RT "Multiple hexose transporters of Schizosaccharomyces pombe.";
RL J. Bacteriol. 182:2153-2162 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

```

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell J.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Symmonds R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Art R., Robben J., Glymonprez B.,
RA Weijens I., Vanstreele E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Usery D., Bartell B.G., Nurse P.,
RA "The genome sequence of Schizosaccharomyces pombe";
RT Nature 415:871-880(2002).
RN [3].
RP SEQUENCE OF 176-546 FROM N.A.
RC STRAIN=PR745;
RX MEDLINE=98162722; PubMed=9501991;
RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT "Identification of open reading frames in Schizosaccharomyces pombe
cDNAs";
RL DNA Res. 4:363-369(1997).
CC -1- FUNCTION: HIGH-AFFINITY GLUCOSE TRANSPORTER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Belongs to the sugar transporter family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF051141; AAC63977.1; -;
DR EMBL; AL031764; CAA21118.1; -;
DR EMBL; D89179; BA013841.1; -;
DR FIC; 140888; T40888;
DR GeneDB Spombe; SPCC1235.14; -;
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub transporter.
DR InterPro; IPR005829; Sug transporter.
DR Pfam; PF00083; Sugar tr_1.
DR PRINTS; PR00171; SUGTRANSPORT.
DR TIGRFAMs; TIGR00879; SP; 1.
DR PROSITE; PS00850; MFS; 1.
DR PROSITE; PS00216; SUGAR TRANSPORT 2; FALSE NEG.
DR PROSITE; PS00217; SUGAR TRANSPORT 1; 1.
KW Transmembrane; Sugar transport; Transport; Glycoprotein.
FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 10 30 1 (POTENTIAL).
FT DOMAIN 31 58 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 59 79 2 (POTENTIAL).
FT DOMAIN 80 87 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 88 108 3 (POTENTIAL).
FT DOMAIN 109 112 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 113 133 4 (POTENTIAL).
FT DOMAIN 134 144 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 145 165 5 (POTENTIAL).
FT DOMAIN 166 179 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 180 200 6 (POTENTIAL).
FT DOMAIN 201 266 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 267 285 7 (POTENTIAL).
FT DOMAIN 286 301 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 302 322 8 (POTENTIAL).

FT DOMAIN 323 328 9 (POTENTIAL).
FT TRANSMEM 329 349 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 350 363 10 (POTENTIAL).
FT TRANSMEM 364 384 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 385 404 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 405 425 11 (POTENTIAL).
FT DOMAIN 426 432 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 433 453 12 (POTENTIAL).
FT DOMAIN 454 546 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 357 357 N-LINKED (GLCNAC... (POTENTIAL).
SQ SEQUENCE 546 AA; 60320 MW; D9B4947CFE2889D7 CRC64;
Query Match 39.8%; Score 41; DB 1; Length 546;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 5 GCYVYIGISFLDQA 18
DB 189 GIITFIGISFLPES 202
RESULT 15
C3G_DROME
ID C3G_DROME STANDARD; PRT; 1571 AA.
AC 077086; Q9W3W3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Guanine nucleotide-releasing factor 2 (CRK SH3-binding GMRP).
GN C3G OR CG3126.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND CHARACTERIZATION.
RC TISSUE=Eye imaginal disk; PubMed=9878058;
RX MEDLINE=99094900; PubMed=9878058;
RA Ishimaru S., Gaul U., Hanafusa H.;
RT "Activation of the Drosophila C3G leads to cell fate changes and
RT overproliferation during development, mediated by the RAS-MAPK
RT pathway and RAPI";
RL EMBO J. 18:145-155(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Fabois B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fiesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hoston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X.-M., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith I.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*,"
RL Science 287:2185-2195(2000).
[3]
RN REVISIONS, AND ALTERNATIVE SPLICING.
RP MEDLINE=22426069; PubMed=12537572;
RX MEDLINE=22426066; PubMed=12537569;
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celnik S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.B.;
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
[4]
RN SEQUENCE OF 1036-1571 FROM N.A.
RP STRAIN=Berkely; TISSUE=Ovary;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celnik S.E.;
RT "A *Drosophila* full-length cDNA resource";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC -!- FUNCTION: Guanine nucleotide-releasing protein that binds to SH3
CC domain of CRK. Transduces signals from CRK to activate RAS. Also
CC involved in MAPK activation.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Names=Long;
CC IsoId=O77086-1; Sequence=D7086-1;
CC Names=Short;
CC IsoId=O77086-2; Sequence=VSP_001823;
CC -!- TISSUE SPECIFICITY: Ubiquitous
CC -!- DEVELOPMENTAL STAGE: Throughout development.
CC -!- SIMILARITY: Contains 1 N-terminal Ras-GEF domain.
CC -!- SIMILARITY: Contains 1 Ras-GEF domain.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 39.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to a
CC frameshift in position 1366.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF053358; AAC35280.1; ALT_FRAME.
CC EMBL; AE003438; AAF46200.2; -.
CC EMBL; AY113355; AAM29360.1; ALT_FRAME.
CC FlyBase; FBgn026145; C3G.
CC DR GO; GO:0005088; F:Ras guanyl-nucleotide exchange factor activity; IMP.
CC DR GO; GO:0007265; P:RAS protein signal transduction; IMP.
CC InterPro; IPR008937; Ras GEF.
CC InterPro; IPR000651; RasGEF.
CC InterPro; IPR001895; RasGEF.
CC Pfam; PF00617; RasGEF; 1.
CC Pfam; PF00618; RasGEFN; 1.
CC SMART; SM00147; RasGEF; 1.

DR SMART; SMO0229; RasGEFN; 1.
DR PROSITE; PS00720; RasGEF; 1.
DR PROSITE; PS50003; RasGEF_CAT; 1.
DR PROSITE; PS50212; RasGEF_NTER; 1.
KW Guanine-nucleotide releasing factor; SH3-binding;
KW Developmental protein; Alternative splicing.
FT DOMAIN 49 69 HIS-RICH.
FT DOMAIN 186 193 POLY-GLY.
FT DOMAIN 195 198 POLY-SER.
FT DOMAIN 213 216 POLY-PRO.
FT DOMAIN 238 247 POLY-GLY.
FT DOMAIN 639 644 POLY-GLN.
FT DOMAIN 699 702 POLY-ALA.
FT DOMAIN 735 802 SER-RICH.
FT DOMAIN 787 791 POLY-THR.
FT DOMAIN 806 814 POLY-GLN.
FT DOMAIN 815 818 POLY-THR.
FT DOMAIN 1170 1292 N-TERMINAL RAS-GEF.
FT DOMAIN 1462 1465 POLY-SER.
FT SITE 546 556 SH3-BINDING (POTENTIAL).
FT SITE 820 831 SH3-BINDING (POTENTIAL).
FT SITE 924 935 SH3-BINDING (POTENTIAL).
FT SITE 986 997 SH3-BINDING (POTENTIAL).
FT VARSPLIC 668 718 Missing (in isoform Short).
FT CONFLICT 19 19 H->Y (IN REF. 1).
FT CONFLICT 93 93 N->S (IN REF. 1).
FT CONFLICT 154 154 E->V (IN REF. 1).
FT CONFLICT 202 202 I->T (IN REF. 1).
FT CONFLICT 243 243 G->GAG (IN REF. 1).
FT CONFLICT 450 453 RYSG -> HYRR (IN REF. 1).
FT CONFLICT 477 477 V->I (IN REF. 1).
FT CONFLICT 807 807 L->Q (IN REF. 1).
FT CONFLICT 818 818 T->TT (IN REF. 1).
FT CONFLICT 851 851 V->A (IN REF. 1).
FT CONFLICT 1004 1004 S->N (IN REF. 1).
FT CONFLICT 1092 1092 L->M (IN REF. 2).
FT CONFLICT 1127 1127 P->A (IN REF. 2).
FT CONFLICT 1311 1311 G->S (IN REF. 2).
SQ SEQUENCE 1571 AA; 172049 MW; B9EF9E8670457684 CRC64;
Query Match 39.3%; Score 40.5; DB 1; Length 1571;
Best Local Similarity 52.9%; Pred.No.1.2e+02;
Matches 9; Conservative 2; Mismatches 5; Indels 1; Gaps 1;
QY 1 AET-PGCVAYIGISFLD 16
Db 1474 AETPPCIPYIGLIQD 1490
Search completed: March 10, 2004, 12:06:31
Job time : 7.15385 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:57:36 ; Search time 31.6667 Seconds
(without alignments)
199.275 Million cell updates/sec

Title: US-10-044-703-52

Perfect score: 103

Sequence: 1 AETPGCVAYIGISFLDQASQ 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	100.0	374	16 Q7U0Z8	Q7U0Z8 mycobacteri
2	91	88.3	374	2 Q49590	Q49590 mycobacteri
3	62	60.2	374	2 Q49589	Q49589 mycobacteri
4	58	56.3	194	2 Q8KUS2	Q8KUS2 enterococcu
5	58	56.3	194	2 Q8KUS0	Q8KUS0 enterococcu
6	58	56.3	194	2 Q8KUS3	Q8KUS3 enterococcu
7	58	56.3	194	2 Q8KIB9	Q8KIB9 enterococcu
8	58	56.3	194	2 Q8KUS1	Q8KUS1 enterococcu
9	58	56.3	194	2 Q8KIB6	Q8KIB6 enterococcu
10	58	56.3	194	2 Q8KH63	Q8KH63 enterococcu
11	58	56.3	194	2 Q8KIC0	Q8KIC0 enterococcu
12	58	56.3	194	2 Q8KUR8	Q8KUR8 enterococcu
13	57	55.3	291	16 Q8SVL1	Q8SVL1 lactobacill
14	53	51.5	302	17 Q8PVG0	Q8PVG0 methanosarc
15	52	50.5	194	2 Q8KUR9	Q8KUR9 enterococcu
16	52	50.5	302	16 Q927Z4	Q927Z4 listeria in

ALIGNMENTS

RESULT 1

Q7U0Z8 AC Q7U0Z8 PRELIMINARY; PRT; 374 AA.
 DT 01-OCT-2003 (TREMBLrel. 25, Created)
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Periplasmic phosphate-binding lipoprotein PstS1 (PBP-1) (PSTS1).
 GN PstS1 OR M50959.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248337; CAD93820.1; -.
 KW Complete proteome.
 SQ SEQUENCE 374 AA; 38215 MW; 1434968191FF201D CRC64;

Query Match 100.0%; Score 103; DB 16; Length 374;
 Best Local Similarity 100.0%; Pred. No. 5.5e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDQASQ 20
 |||||
 Db 239 AETPGCVAYIGISFLDQASQ 258

RESULT 2

Q49590 PRELIMINARY; PRT; 374 AA.
 ID Q49590

Q8Y4E7 listeria mo
 Q834F0 enterococcu
 Q9CEW3 lactococcu
 Q8XPK4 raiistonias
 Q8KZ9 chlorobium
 Q8DZV4 streptococ
 Q8DU20 streptococ
 Q8ESK2 streptococ
 Q878P6 streptococ
 Q99ZG1 streptococ
 Q8P0V0 streptococ
 Q8K7C3 streptococ
 Q8R9H9 thermosarc
 Q8P2B6 methanosarc
 Q9VE77 drosophila
 Q7WH75 bordetella
 Q7W9I5 bordetella
 Q7VWP2 bordetella
 Q8KZ8 chlorobium
 Q04622 arabidopsis
 Q8G6E9 bifidobacte
 Q9Fdk9 zymomonas m
 Q92WY5 rhizobium m
 Q7Y1B7 hordeum vul
 Q8P52 xanthomonas
 Q928B7 chlamydia p
 Q97C31 streptococ
 Q8PBI1 streptococ
 Q9PHD0 xyella fas

```
AC Q49590;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PstS subunit of ABC transporter.
GN PstS1B.
OS Mycobacterium intracellulare.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1767;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35761;
RX MEDLINE=97055782; PubMed=8900068;
RA Thangaraj H.S., Bull T.J., De Smet K.A.L., Hill M., Rouse D.A.,
RA Moreno C., Ivanyi J.;
RA "Duplication of genes encoding the immunodominant 38 kDa antigen in
RT Mycobacterium intracellulare.";
RL FEMS Microbiol. Lett. 144:235-240(1996).
DR EMBL; X95538; CAA64784.1; -
DR HSP; P06128; 1FXG
DR GO; GO:000515; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000437; Prok_lipoport_S.
DR InterPro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
SQ SEQUENCE 374 AA; 38068 MW; A7FD25843660B9D7 CRC64;

Query Match 88.3%; Score 91; DB 2; Length 374;
Best Local Similarity 94.4%; Pred. No. 5.9e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDOA 18
DB 238 ADTEGCVAYIGISFLDOA 255

RESULT 3
Q49589 PRELIMINARY; PRT; 374 AA.
AC Q49589;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PstS subunit of ABC transporter.
GN PstS1A.
OS Mycobacterium intracellulare.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1767;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35761;
RX MEDLINE=97055782; PubMed=8900068;
RA Thangaraj H.S., Bull T.J., De Smet K.A.L., Hill M., Rouse D.A.,
RA Moreno C., Ivanyi J.;
RA "Duplication of genes encoding the immunodominant 38 kDa antigen in
RT Mycobacterium intracellulare.";
RL FEMS Microbiol. Lett. 144:235-240(1996).
DR EMBL; X95538; CAA64783.1; -
DR HSP; P06128; 1A54.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000437; Prok_lipoport_S.
DR InterPro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
SQ SEQUENCE 374 AA; 37998 MW; 45796D4E9F6F513D CRC64;

Query Match 60.2%; Score 62; DB 2; Length 374;
Best Local Similarity 60.0%; Pred. No. 0.048;
Matches 12; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
```

```
QY 1 AETPGCVAYIGISFLDOASQ 20
DB 239 AANFGCVAYTSGSLDRAQ 258

RESULT 4
Q8KUS2 PRELIMINARY; PRT; 194 AA.
AC Q8KUS2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Phosphate ABC transporter (Fragment).
GN PstS.
OS Enterococcus faecium (Streptococcus faecium).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1352;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A013 ST33;
RX MEDLINE=22033280; PubMed=12037049;
RA Homan W.L., Tribe D., Poznanski S., Li M., Hogg G., Spalburg E.,
RA Van Embden J.D.A., Willems R.J.L.;
RT "Multilocus Sequence Typing Scheme for Enterococcus faecium.";
RL J. Clin. Microbiol. 40:1963-1971(2002).
DR EMBL; AF443372; AAM49692.1; -
DR NON_TER 1
DR NON_TER 194
SQ SEQUENCE 194 AA; 21305 MW; 11CE49DDB3820P1F CRC64;

Query Match 56.3%; Score 58; DB 2; Length 194;
Best Local Similarity 40.0%; Pred. No. 0.12;
Matches 8; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDOASQ 20
DB 112 SOTPGAISYLAFLDSTQ 131

RESULT 5
Q8KUS0 PRELIMINARY; PRT; 194 AA.
AC Q8KUS0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Phosphate ABC transporter (fragment).
GN PstS.
OS Enterococcus faecium (Streptococcus faecium).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1352;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ned D0154;
RX MEDLINE=22033280; PubMed=12037049;
RA Homan W.L., Tribe D., Poznanski S., Li M., Hogg G., Spalburg E.,
RA Van Embden J.D.A., Willems R.J.L.;
RT "Multilocus Sequence Typing Scheme for Enterococcus faecium.";
RL J. Clin. Microbiol. 40:1963-1971(2002).
DR EMBL; AF443376; AAM49696.1; -
DR NON_TER 1
DR NON_TER 194
SQ SEQUENCE 194 AA; 21275 MW; 164A9C641B769E3A CRC64;

Query Match 56.3%; Score 58; DB 2; Length 194;
Best Local Similarity 40.0%; Pred. No. 0.12;
Matches 8; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 AETPGCVAYIGISFLDOASQ 20
DB 112 SOTPGAISYLAFLDSTQ 131
```


DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Phosphate ABC transporter (Fragment).
 GN PSTs.
 OS Enterococcus faecium (Streptococcus faecium).
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
 OX NCBI_TaxID=1352;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=E0072 ST19, and Ne D0161;
 RX MEDLINE=22033280; PubMed=12037049;
 RA Homan W.L., Tribe D., Poznanski S., Li M., Hogg G., Spalburg E.,
 RA Van Embden J.D.A., Willems R.J.L.;
 RT "Multilocus Sequence Typing Scheme for Enterococcus faecium.";
 RL J. Clin. Microbiol. 40:1963-1971(2002).
 DR EMBL; AF443378; AAM49698.1; -;
 DR EMBL; AF443380; AAM49700.1; -;
 FT NON TER 1 194
 FT NON TER 1 194
 SQ SEQUENCE 194 AA; 21275 MW; E755946B2FC01D9C CRC64;
 Query Match 56.3%; Score 58; DB 2; Length 194;
 Best Local Similarity 40.0%; Pred. No. 0.12;
 Matches 8; Conservative 8; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 AETPGCVAYIGISFLDQASQ 20
 Db 112 SOTPGAISYLAFLSYLDSTQ 131
 RESULT 11
 Q8KIC0 PRELIMINARY; PRT; 194 AA.
 ID Q8KIC0
 AC Q8KIC0
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Phosphate ABC transporter (Fragment).
 GN PSTs.
 OS Enterococcus faecium (Streptococcus faecium).
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
 OX NCBI_TaxID=1352;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=E0158 ST16, A012 ST30, and Ne D0166;
 RX MEDLINE=22033280; PubMed=12037049;
 RA Homan W.L., Tribe D., Poznanski S., Li M., Hogg G., Spalburg E.,
 RA Van Embden J.D.A., Willems R.J.L.;
 RT "Multilocus Sequence Typing Scheme for Enterococcus faecium.";
 RL J. Clin. Microbiol. 40:1963-1971(2002).
 DR EMBL; AF443368; AAM49688.1; -;
 DR EMBL; AF443369; AAM49689.1; -;
 DR EMBL; AF443381; AAM49701.1; -;
 FT NON TER 1 194
 FT NON TER 1 194
 SQ SEQUENCE 194 AA; 21319 MW; 043BA9DDB398151F CRC64;
 Query Match 56.3%; Score 58; DB 2; Length 194;
 Best Local Similarity 40.0%; Pred. No. 0.12;
 Matches 8; Conservative 8; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 AETPGCVAYIGISFLDQASQ 20
 Db 112 SOTPGAISYLAFLSYLDSTQ 131
 RESULT 12
 Q8KUR8 PRELIMINARY; PRT; 194 AA.
 ID Q8KUR8
 AC Q8KUR8
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Phosphate ABC transporter (Fragment).
 GN PSTs.
 OS Enterococcus faecium (Streptococcus faecium).
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
 OX NCBI_TaxID=1352;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ne D0307;
 RX MEDLINE=22033280; PubMed=12037049;
 RA Homan W.L., Tribe D., Poznanski S., Li M., Hogg G., Spalburg E.,
 RA Van Embden J.D.A., Willems R.J.L.;
 RT "Multilocus Sequence Typing Scheme for Enterococcus faecium.";
 RL J. Clin. Microbiol. 40:1963-1971(2002).
 DR EMBL; AF443382; AAM49702.1; -;
 FT NON TER 1 194
 FT NON TER 1 194
 SQ SEQUENCE 194 AA; 21341 MW; BB7BA2290BEC6A9F CRC64;
 Query Match 56.3%; Score 58; DB 2; Length 194;
 Best Local Similarity 40.0%; Pred. No. 0.12;
 Matches 8; Conservative 8; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 AETPGCVAYIGISFLDQASQ 20
 Db 112 SOTPGAISYLAFLSYLDSTQ 131
 RESULT 13
 Q88YL1 PRELIMINARY; PRT; 291 AA.
 ID Q88YL1
 AC Q88YL1
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Phosphate ABC transporter, substrate binding protein.
 GN PSTE OR LP 0746.
 OS Lactobacillus plantarum.
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
 OC Lactobacillus.
 OX NCBI_TaxID=1590;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIMB 8826 / WCFS1;
 RX MEDLINE=22480296; PubMed=12566566;
 RA Kleschevskiy M., Boekhorst J., van Kranenburg R., Molenaar D.,
 RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
 RA Fiers M.W.B.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
 RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
 RA De Vos W.M., Siezen R.J.;
 RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
 DR EMBL; AL935254; CAD63340.1; -;
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR006059; SBP_bac_1.
 DR Pfam; PF01547; SBP_bac_1; 1.
 KW Complete proteome.
 SQ SEQUENCE 291 AA; 31528 MW; 9587CB29D78B67DE CRC64;
 Query Match 55.3%; Score 57; DB 16; Length 291;
 Best Local Similarity 45.0%; Pred. No. 0.26;
 Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
 Qy 1 AETPGCVAYIGISFLDQASQ 20
 Db 193 ASTPGAISYVAFLSYLDNSVQ 212
 RESULT 14
 Q8PVG0 PRELIMINARY; PRT; 302 AA.
 ID Q8PVG0
 AC Q8PVG0
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)

Db 112 SQTGAI SYLA FSNLDDSTQ 131

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:50:51 ; Search time 47.5641 Seconds
(without alignments)
118.807 Million cell updates/sec

Title: US-10-044-703-61

Perfect score: 97

Sequence: 1 NFGATIEAAASAIQGNVTSIH 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97	100.0	20	4	Aae12276 Mycobacte
2	97	100.0	51	2	Aaw32339 Mycobacte
3	97	100.0	51	2	Aaw32456 Mycobacte
4	97	100.0	51	2	Aaw64334 Mycobacte
5	97	100.0	51	2	Aaw81701 M. tuberc
6	97	100.0	51	2	Aay38988 M. tuberc
7	97	100.0	51	2	Aay39131 M. tuberc
8	97	100.0	51	4	Aau01897 Mycobacte
9	97	100.0	51	5	Aae29716 Mycobacte
10	97	100.0	51	5	Aae17580 Mycobacte
11	97	100.0	95	2	Aaw11494 ESAT-6.3/
12	97	100.0	95	2	Aay29788 Mycobacte
13	97	100.0	95	2	Aay29890 Mycobacte
14	97	100.0	95	4	Aab35219 M. tubercu
15	97	100.0	95	5	ABG30968 Mycobacte
16	97	100.0	95	5	Aam50740 Mycobacte
17	97	100.0	95	7	Ade28550 Mycobacte
18	97	100.0	142	7	Ade28541 Human/Myc
19	97	100.0	145	7	Ade28543 Human/Myc
20	97	100.0	146	7	Ade28545 Human/Myc
21	97	100.0	146	7	Ade28547 Human/Myc
22	97	100.0	196	2	Aay29888 Mycobacte
23	97	100.0	403	2	Aaw72943 Mycobacte
24	97	100.0	403	2	Aay21963 Amino aci
25	97	100.0	404	2	Aaw72942 Mycobacte

26	97	100.0	404	2	AAY21962	Aay21962 Amino aci
27	85	87.6	20	4	AAE12275	Aae12275 Mycobacte
28	67	69.1	15	3	AAAY94599	Aay94599 Mycobacte
29	67	69.1	15	5	ABG30953	Abg30953 Mycobacte
30	66	68.0	15	3	AAAY94595	Aay94595 Mycobacte
31	66	68.0	15	5	ABG30952	Abg30952 Mycobacte
32	66	68.0	15	5	AAO17433	Aao17433 Early sec
33	60	61.9	13	4	AAE12274	Aae12274 Mycobacte
34	60	61.9	19	2	AAW35548	Aaw35548 ESAT-6 an
35	55	56.7	455	2	AAW41076	Aaw41076 P. furios
36	54	55.7	15	3	AAAY94591	Aay94591 Mycobacte
37	54	55.7	15	5	ABG30954	Abg30954 Mycobacte
38	51	52.6	17	3	AAAY8581	Aay8581 Antigenic
39	47	48.5	540	5	ABB54282	Abb54282 Lactococc
40	46	47.4	20	4	AAE12277	Aae12277 Mycobacte
41	46	47.4	170	2	AAW06939	Aaw06939 Cagl locu
42	46	47.4	682	3	AAAY1947	Aay1947 A. sydcwl
43	44.5	45.9	150	6	ABU23581	Abu23581 Protein e
44	44	45.4	307	6	ABU27060	Abu27060 Protein e
45	44	45.4	319	2	AAY37655	Aay37655 Amino aci

ALIGNMENTS

RESULT 1
AAE12276
ID AAE12276 standard; peptide; 20 AA.
XX AAE12276;
AC AAE12276;
XX
DT 18-DEC-2001 (first entry)
XX
DE Mycobacterium tuberculosis (Mtb) peptide #61.
XX
KW Mycobacterium tuberculosis; Mtb peptide; antibacterial; vaccine;
KW infection; anti-Mtb immune response.
XX
OS Mycobacterium tuberculosis.
XX
FN WO200170774-A2.
XX
PD 27-SEP-2001.
XX
PF 20-MAR-2001; 2001WO-US008906.
XX
PR 20-MAR-2000; 2000US-0190834P.
XX
(UYBR-) UNIV BROWN RES FOUND.
XX
PA Degroot AS;
XX
DR WPI; 2001-616401/71.
XX
PT New vaccine for immunizing a mammalian subject, preferably humans,
PT against infection caused by Mycobacterium tuberculosis.
XX
PS Disclosure; Fig 4; 42pp; English.
XX
CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine
CC candidate peptides. The invention also relates to a method for
CC identifying Mtb vaccine candidate peptides as well as vaccines comprising
CC these candidate peptides. Vaccines of the invention and Mtb vaccine
CC candidate peptides are useful for inducing an anti- Mycobacterium
CC tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a
CC mammalian subject preferably human. They are used for immunising a
CC mammalian subject, preferably humans, against infection caused by
CC Mycobacterium tuberculosis. The present sequence is a Mtb vaccine
CC candidate peptide
SQ Sequence 20 AA;
Query Match 100.0%; Score 97; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.1e-09; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0

QY 1 NFAGIEAAASAIQGNVTSIH 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 NFAGIEAAASAIQGNVTSIH 20

RESULT 2
AAW32339
ID AAW32339 standard; protein; 51 AA.
XX AC
XX AAW32339;
XX 13-JAN-1998 (first entry)
XX XX
XX Mycobacterium tuberculosis antigen ESAT-6.
DE XX
XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX OS
XX Mycobacterium tuberculosis.
XX PN
XX WO9709429-A2.
XX PD
XX 13-MAR-1997.
XX PF
XX 30-AUG-1996; 96WO-US014675.
XX PR
XX 01-SEP-1995; 95US-00523435.
XX PR 22-SEP-1995; 95US-00532136.
XX PR 22-MAR-1996; 96US-00620280.
XX PR 05-JUN-1996; 96US-00658800.
XX PR 12-JUL-1996; 96US-00680573.
XX PA
XX (CORI-) CORIXA CORP.
XX PI
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TH, Twardzik DR;
XX DR
XX WPI; 1997-192904/17.
XX DR N-PSDB; AAT91463.
XX XX
XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
PT useful for diagnosis of M. tuberculosis infection.
XX PS
XX Claim 43; Page 147; 190pp; English.
XX CC
XX A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
CC variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a specifically claimed
CC sequence of the ESAT-6 M.tuberculosis antigen. The immunogenic
CC polypeptide can be used to diagnose M.tuberculosis infection by forming
CC complexes with specific antibodies in the sample. Fragments of DNA
CC encoding the immunogenic polypeptide can be used as diagnostic primers or
CC probes and agents that bind to the antigen, especially monoclonal
CC antibodies or equivalent polyclonal antibodies, are also used for
XX diagnosis
XX SQ
XX Sequence 51 AA;

Query Match 100.0%; Score 97; DB 2; Length 51;
Best Local Similarity 100.0%; Pred. No. 3.4e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAAASAIQGNVTSIH 20
| | | | | | | | | | | | | | | | | | | | | |
Db 7 NFAGIEAAASAIQGNVTSIH 26

RESULT 3
AAW32466

ID AAW32466 standard; protein; 51 AA.
XX AC
XX AAW32466;
XX 09-JAN-1998 (first entry)
XX DE
XX Mycobacterium tuberculosis antigen ESAT-6.
XX KW
XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX XX
XX Mycobacterium tuberculosis.
XX PN
XX WO9709428-A2.
XX PD
XX 13-MAR-1997.
XX PF
XX 30-AUG-1996; 96WO-US014674.
XX PR
XX 01-SEP-1995; 95US-00523436.
XX PR 22-SEP-1995; 95US-00533634.
XX PR 22-MAR-1996; 96US-00620874.
XX PR 05-JUN-1996; 96US-00659683.
XX PR 12-JUL-1996; 96US-00680574.
XX PA
XX (CORI-) CORIXA CORP.
XX PI
XX Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TH, Twardzik DR;
XX DR
XX WPI; 1997-192903/17.
XX DR N-PSDB; AAT91529.
XX XX
XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
PT useful in vaccines for prevention or treatment of tuberculosis, also for
PT diagnosis.
XX PS
XX Disclosure; Page 135; 168pp; English.
XX CC
XX A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
CC variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M.tuberculosis antigen,
CC ESAT-6. The immunogenic protein, and fusion proteins containing one or
CC more of the proteins or one of the proteins plus ESAT-6, are useful in
CC vaccines, preferably when formulated with a non-specific adjuvant, to
CC induce an immune response against M.tuberculosis (for treatment or
CC prevention)
XX SQ
XX Sequence 51 AA;

Query Match 100.0%; Score 97; DB 2; Length 51;
Best Local Similarity 100.0%; Pred. No. 3.4e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAAASAIQGNVTSIH 20
| | | | | | | | | | | | | | | | | | | | | |
Db 7 NFAGIEAAASAIQGNVTSIH 26

RESULT 4
AAW64334
ID AAW64334 standard; protein; 51 AA.
XX AC
XX AAW64334;
XX 09-NOV-1998 (first entry)
XX XX
XX Mycobacterium tuberculosis antigen ESAT-6.
DE XX
XX Tuberculosis; infection; diagnosis; antigen; ESAT-6.
XX OS
XX Mycobacterium tuberculosis.

XX WO9816645-A2.
 PN
 XX 23-APR-1998.
 PD
 XX 07-OCT-1997; 97WO-US018214.
 XX
 XX 11-OCT-1996; 96US-00729622.
 PR
 PR 13-MAR-1997; 97US-00818111.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX
 XX WPI; 1998-251292/22.
 DR
 DR N-PSDB; AAV44393.
 XX
 XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and
 PT diagnosis of tuberculosis.
 XX
 XX Disclosure; Page 130; 250pp; English.
 PS
 XX This polypeptide comprises an antigenic portion of Mycobacterium
 CC tuberculosis antigen ESAT-6. The invention relates to compositions and
 CC methods for diagnosing tuberculosis. It provides polypeptides (see
 CC AAW64291-W64379) comprising an antigenic portion of a soluble M.
 CC tuberculosis antigen, or an immunogenic portion of an M. tuberculosis
 CC antigen, as well as fusion proteins between these polypeptides and known
 CC antigens such as ESAT-6. Also claimed are methods and diagnostic kits for
 CC detecting M. tuberculosis infection in a patient using these
 CC polypeptides, antibodies, or oligonucleotide probes and primers
 XX
 XX Sequence 51 AA;
 SQ

Query Match 100.0%; Score 97; DB 2; Length 51;
 Best Local Similarity 100.0%; Pred. No. 3.4e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NFAGIEAAASAIQGNVTSIH 20
 DB 7 NFAGIEAAASAIQGNVTSIH 26

RESULT 5
 ID AAW81701 standard; protein; 51 AA.
 AC AAW81701;
 XX
 XX 27-JAN-1999 (first entry)
 DT
 DE M. tuberculosis immunogenic polypeptide ESAT-6.
 XX
 XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 KW vaccine; pharmaceutical; infection; diagnosis.
 XX
 XX Mycobacterium tuberculosis.
 OS
 XX WO9816646-A2.
 PN
 XX 23-APR-1998.
 PD
 XX 07-OCT-1997; 97WO-US018293.
 XX
 XX 11-OCT-1996; 96US-00730510.
 PR
 PR 13-MAR-1997; 97US-00818112.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;

XX WPI; 1998-261042/23.
 DR
 DR N-PSDB; AAV64501.
 XX
 XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and for
 PT diagnosis, treatment and prevention of tuberculosis.
 XX
 XX Disclosure; Page 126; 230pp; English.
 PS
 XX This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
 CC inducing protective immunity against tuberculosis (TB). This sequence can
 CC be formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis
 XX
 XX Sequence 51 AA;
 SQ

Query Match 100.0%; Score 97; DB 2; Length 51;
 Best Local Similarity 100.0%; Pred. No. 3.4e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NFAGIEAAASAIQGNVTSIH 20
 DB 7 NFAGIEAAASAIQGNVTSIH 26

RESULT 6
 ID AAY38988 standard; protein; 51 AA.
 AC AAY38988;
 XX
 XX 05-NOV-1999 (first entry)
 DT
 DE M. tuberculosis recombinant antigen protein ESAT-6.
 XX
 XX Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 XX
 XX Mycobacterium tuberculosis.
 OS
 XX WO9942118-A2.
 PN
 XX 26-AUG-1999.
 PD
 XX 17-FEB-1999; 99WO-US003265.
 PF
 XX 18-FEB-1998; 98US-00024753.
 PR
 PR 05-MAY-1998; 98US-00072596.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX
 XX WPI; 1999-527416/44.
 DR
 DR N-PSDB; AAZ19091.
 XX
 XX New polypeptide comprising antigenic portions of M. tuberculosis.
 PT
 XX Claim 51; Page 165; 323pp; English.
 PS
 XX This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against M.
 CC tuberculosis infection. The new detection methods are needed as current
 CC vaccination strategies do not provide 100% immunity
 XX
 XX Sequence 51 AA;
 SQ

Query Match 100.0%; Score 97; DB 2; Length 51;
 Best Local Similarity 100.0%; Pred. No. 3.4e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFAGIEAASAIQGNVTSIH 20
 DB 7 NFAGIEAASAIQGNVTSIH 26

RESULT 7

AAI39131
 ID AAY39131 standard; protein; 51 AA.

XX AC AAY39131;

XX XX 05-NOV-1999 (first entry)

XX DE M. tuberculosis ESAT-6 amino acid sequence.

XX XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.

XX OS Mycobacterium tuberculosis.

XX PN WO9942076-A2.

XX XX 26-AUG-1999.

XX PF 17-FEB-1999; 99WO-US003268.

XX PR 18-FEB-1998; 98US-00025197.

XX PR 03-MAY-1998; 98US-00072967.

XX XX (CORI-) CORIXA CORP.

XX PI Read SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;

XX DR WPI; 1999-527409/44.

XX DR N-PSDB; AAZ19303.

XX XX New antigens from Mycobacterium tuberculosis useful in diagnostic skin
 tests and protective or therapeutic vaccines or compositions.

XX PS Disclosure; Page 120; 299pp; English.

XX CC The present invention describes polypeptides comprising an immunogenic
 part of a Mycobacterium tuberculosis antigen (Ag). Also described are
 CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
 CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
 CC polypeptides fragments, can be used in pharmaceutical compositions or
 CC vaccines to generate a protective or therapeutic immune response to M.
 CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
 CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
 CC killer cells and/or macrophages in tuberculosis-immune subjects. AAZ19249
 CC to AAZ19460 and AAZ19083 to AAZ19225 are used in the exemplification of
 CC the present invention

XX XX Sequence 51 AA;

Query Match 100.0%; Score 97; DB 2; Length 51;
 Best Local Similarity 100.0%; Pred. No. 3.4e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFAGIEAASAIQGNVTSIH 20

DB 7 NFAGIEAASAIQGNVTSIH 26

RESULT 8

AAU01897

ID AAU01897 standard; protein; 51 AA.
 XX AC AAU01897;
 XX DT 29-AUG-2001 (first entry)
 XX XX Mycobacterium tuberculosis partial antigen ESAT-6.
 DE ESAT-6; antigen; vaccine; tuberculosis; AIDS;
 XX KW acquired immunodeficiency disease.
 XX OS Mycobacterium tuberculosis.
 XX PN WO200124820-A1.
 XX PD 12-APR-2001.
 XX PF 10-OCT-2000; 2000WO-US028095.
 XX PR 07-OCT-1999; 99US-0158338P.
 XX PR 07-OCT-1999; 99US-0158425P.
 XX PA (CORI-) CORIXA CORP.
 XX PI Skeiky Y, Reed S, Houghton RL, Mcneill PD, Dillon DC, Lodes ML;
 XX DR WPI; 2001-290576/30.
 XX DR N-PSDB; AAS03787.
 XX PT Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens.
 XX PS Disclosure; Page 164; 168pp; English.
 XX CC The sequence represents Mycobacterium tuberculosis ESAT-6 antigen.
 CC Compositions comprising at least 2 heterologous antigens, as a fusion
 CC protein, and vectors expressing the fusion proteins are used as vaccines
 CC to prophylactically immunise mammals (especially humans) against
 CC infection by Mycobacteria. The compositions contain at least 2
 CC heterologous antigens that increase the serological sensitivity of
 CC individuals infected with tuberculosis, a disease frequently affecting
 CC patients with acquired immunodeficiency disease, AIDS
 XX XX Sequence 51 AA;
 Query Match 100.0%; Score 97; DB 4; Length 51;
 Best Local Similarity 100.0%; Pred. No. 3.4e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 NFAGIEAASAIQGNVTSIH 20
 DB 7 NFAGIEAASAIQGNVTSIH 26
 RESULT 9
 AAE29716
 ID AAE29716 standard; protein; 51 AA.
 XX AC AAE29716;
 XX XX 27-JAN-2003 (first entry)
 XX DE Mycobacterium tuberculosis ESAT-6 antigenic protein.
 XX KW Vaccine; immunity; diagnostic agent; gene therapy; ESAT-6 antigen.
 XX OS Mycobacterium tuberculosis.
 XX PN WO200272792-A2.
 XX PD 19-SEP-2002.
 XX PF 13-MAR-2002; 2002WO-US008223.

XX 13-MAR-2001; 2001US-0275837P.
 XX (CORI-) CORIXA CORP.
 XX
 XX Skeiky Y, Brannon M, Guderian J;
 PI WPI; 2002-759844/82.
 XX N-PSDB; AAD29716.
 DR
 XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.
 XX
 XX Disclosure; Page 106; 155pp; English.
 PS
 XX The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is M.
 CC tuberculosis ESAT-6 antigenic protein
 XX
 XX Sequence 51 AA;
 SQ

Query Match 100.0%; Score 97; DB 5; Length 51;
 Best Local Similarity 100.0%; Pred. NO. 3.4e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFAGIEAAAGAIQGNVTSIH 20
 ID AAE17580 standard; protein; 51 AA.
 XX AAE17580;
 AC AAE17580;
 XX
 XX 22-APR-2002 (first entry)
 DT Mycobacterium species ESAT-6 protein.
 DE
 XX Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; ESAT-6 protein.
 XX Mycobacterium sp.
 OS WO200198460-A2.
 XX
 XX 27-DEC-2001.
 PD
 XX 20-JUN-2001; 2001WO-US019959.
 PF
 XX 20-JUN-2000; 2000US-00597796.
 PR 01-FEB-2001; 2001US-0265737P.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX Skeiky Y, Reed S, Alderson M;
 PI WPI; 2002-147798/19.
 XX N-PSDB; AAD28351.
 DR
 XX

PT Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 XX Claim 9; Page 122; 136pp; English.
 XX
 XX The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC ESAT-6 protein
 XX
 XX Sequence 51 AA;
 SQ

Query Match 100.0%; Score 97; DB 5; Length 51;
 Best Local Similarity 100.0%; Pred. NO. 3.4e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFAGIEAAAGAIQGNVTSIH 20
 ID AAW11494 standard; protein; 95 AA.
 XX AAW11494;
 AC AAW11494;
 XX
 XX 27-MAR-1997 (first entry)
 DT ESAT6.
 DE
 XX ESAT6; Mycobacterium tuberculosis; 6kDa antigen; tuberculosis complex;
 KW HYB76-8 reactive antigen; interferon-gamma; memory T-lymphocyte; vaccine;
 KW M. africanum; M. bovis; delayed-type hypersensitivity reaction; therapy.
 XX
 XX Mycobacterium tuberculosis.
 OS WO9501441-A1.
 XX
 XX 12-JAN-1995.
 PD
 XX 01-JUL-1994; 94WO-DK000273.
 PF
 XX 02-JUL-1993; 93DK-00000798.
 PR
 XX (STAT-) STATENS SERUMSINSTITUT.
 PA
 XX Andersen P, Andersen AB, Haslov K, Sorensen A;
 PI WPI; 1995-061005/08.
 XX N-PSDB; AAT51422.
 DR
 XX Vaccine for tuberculosis induces interferon-gamma release from T-
 PT lymphocytes - comprises an antigen released from mycobacteria, for
 PT immunisation of humans.
 PT
 XX

PS Claim 5; Page 61-63; 101pp; English.

XX This sequence represents the Mycobacterium tuberculosis ESAT6. ESAT6 is
CC also known as the 6kDa antigen, or the HVB76-8 reactive antigen. ESAT6 is
CC released from metabolising bacteria, and can be isolated from short term
CC filtrates grown as shaken cultures for 7 days. ESAT6 also induces a
CC release of interferon-gamma from reactivated memory T-lymphocytes. This
CC protein sequence is included in the vaccine of the invention. The vaccine
CC is for immunising an animal (including humans) against tuberculosis
CC caused by a Mycobacteria belonging to the tuberculosis complex. The
CC Mycobacteria of the tuberculosis complex are, M. tuberculosis, M.
CC africanum, and M. bovis. The vaccine can evoke a protective immune
CC response against tuberculosis or a delayed-type hypersensitivity
CC reaction. The protein can also be included in a composition for
CC diagnosing tuberculosis. The composition is injected intradermally, and a
CC skin reaction is an indicator of tuberculosis
XX
SQ Sequence 95 AA;

Query Match 100.0%; Score 97; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 7.3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAAASAIQGNVTSIH 20
DB 7 NFAGIEAAASAIQGNVTSIH 26

RESULT 12

AAAY29788
ID AAY29788 standard; protein; 95 AA.

AC AAY29788;

DT 08-NOV-1999 (first entry)

DE Mycobacterial tuberculosis ESAT-6 protein.

XX Mycobacterial; ESAT-6; 6 kDa antigen; tuberculosis; infection;
XX immunological response; diagnosis; vaccine.

OS Mycobacterium tuberculosis.

XX US955077-A.

XX 21-SEP-1999.

XX 05-JUN-1995; 95US-00465640.

XX 20-SEP-1993; 93US-00123182.

PR 01-JUL-1994; 94WO-DK000273.

XX (STAT-) STATENS SERUMINSTITUT.

XX Andersen P, Haslov K, Sorensen AL, Andersen AB;

XX WPI; 1999-539545/45.

DR N-PSDB; AAZ08877.

XX Polypeptide secreted from Mycobacterium is useful as a vaccine against
PT tuberculosis.

PS Claim 24; Fig 10C; 39pp; English.

XX The present invention describes a purified or non-naturally occurring
CC polypeptide (I) released from a metabolising mycobacteria comprising an
CC ESAT6, also called the 6 kDa antigen. The present sequence represents
CC Mycobacterium tuberculosis ESAT-6 protein. Also described is a purified
CC or non-naturally occurring polypeptide (II) with the ability to elicit a
CC delayed type hypersensitivity reaction which comprises a T cell epitope
CC of (I). (II) can be used with a carrier or vehicle in a composition for
CC diagnosing tuberculosis caused by mycobacteria belonging to the
CC tuberculosis complex, i.e. Mycobacterium tuberculosis, M. bovis and M.

CC africanum. The composition can be used to detect microbial antibodies or
CC components of mycobacteria in samples or in animals through the use of
CC immunoassays. (II) can be used as a vaccine for immunising an animal,
CC including humans against tuberculosis caused by mycobacteria of the
CC tuberculosis-complex. (II) induce a release of IFN-gamma from reactivated
CC T-lymphocytes evoking a protective immune response. Vaccine containing
CC (II) has the same protective potency as the live BCG vaccine against
CC tuberculosis

SQ Sequence 95 AA;

Query Match 100.0%; Score 97; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 7.3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAAASAIQGNVTSIH 20
DB 7 NFAGIEAAASAIQGNVTSIH 26

RESULT 13

AAAY29890
ID AAY29890 standard; protein; 95 AA.

AC AAY29890;

XX 18-NOV-1999 (first entry)

XX Mycobacterium tuberculosis ESAT-6 protein sequence.

XX Mycobacterial; lactic acid bacterium; diagnosis; skin test; vaccine;
XX delayed type hypersensitivity; DTH; ESAT-6 homodimer; tuberculosis;
XX interferon-gamma release.

OS Mycobacterium tuberculosis.

XX WO9945119-A2.

XX 10-SEP-1999.

XX 05-MAR-1999; 99WO-DK000109.

XX 06-MAR-1998; 98DK-00000306.

PR 06-MAR-1998; 98US-0077105P.

XX (STAT-) STATENS SERUM INST.

XX Jensen CL, Folkersen J;

XX WPI; 1999-551043/46.

DR N-PSDB; AAZ21132.

XX New mycobacterial polypeptide produced in lactic acid bacteria, useful in
PT tuberculosis diagnosis and vaccines.

XX Disclosure; Page 75-76; 76pp; English.

XX The present invention describes a bioactive polypeptide (or
CC immunologically equivalent analogue) produced in lactic acid bacteria
CC which reacts with lymphoid cells primed with Mycobacterium tuberculosis
CC complex mycobacteria (M. tuberculosis, M. africanum or M. bovis). The
CC polypeptide and ESAT-6 polypeptides are useful in compositions for
CC diagnosis of and vaccination against tuberculosis caused by M.
CC tuberculosis complex mycobacteria. The ESAT-6 polypeptide can be used to
CC diagnose ongoing/previous sensitisation with these bacteria by detecting
CC cytokine release when contacting blood samples with the polypeptide. The
CC bioactive polypeptide may be used in diagnostic compositions and
CC vaccines for mycobacteria other than of the M. tuberculosis complex, e.g.
CC M. avium which infects poultry and occasionally humans, M. leprae; they
CC are especially useful when they do not react with lymphoid cells
CC previously primed with M. tuberculosis complex mycobacteria, and so do
CC not give rise to a diagnostic reaction in individuals infected with these
CC bacteria. The polypeptides may also be used in in vitro diagnostic tests

CC e.g. stimulation of interferon-gamma release from lymphocytes. The
 CC polypeptide has similar or higher bioactivity as currently used
 CC tuberculin reagent in the standard delayed type hypersensitivity (DTH)
 CC skin test for tuberculosis, but may have greater specificity, being
 CC better able to discriminate between lymphoid cells primed from
 CC tuberculosis and from previous vaccination. The present sequence
 CC represents M. tuberculosis ESAT-6 used in the exemplification of the
 CC present invention

XX Sequence 95 AA;

Query Match 100.0%; Score 97; DB 2; Length 95;
 Best Local Similarity 100.0%; Pred. No. 7.3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAASAIQGNVTSIH 20
 DB 7 NFAGIEAASAIQGNVTSIH 26

RESULT 14

AAB35219
 ID AAB35219 standard; protein; 95 AA.

XX AC AAB35219;

DT 24-APR-2001 (first entry)

DE M tuberculosis Rv3875 protein.

KW Tuberculosis; TB; vaccine; esat-6 gene family; Rv0287; Rv1036c; Rv1037c;
 KW Rv2346c; Rv2348c; Rv2653c; Rv2654c; Rv3020c; Rv3444c; Rv3445c; Rv3890c;
 KW Rv3891c; Rv3904c; Rv3905c.

XX Mycobacterium tuberculosis.

XX WO200104151-A2.

XX 18-JAN-2001.

XX 13-JUL-2000; 2000WO-DK000398.

PR 13-JUL-1999; 99DK-00001020.

PR 15-JUL-1999; 99US-0144011P.

XX (STAP-) STATENS SERUM INST.

XX Andersen P, Skjot R;

DR WPI; 2001-091923/10.

XX New polypeptide encoded by a member of the esat-6-gene family for
 PT immunizing against and diagnosis of tuberculosis.

XX Example 2; Page 65; 80pp; English.

CC The present invention provides the protein and coding sequences for
 CC members of the esat-6 gene family from Mycobacterium tuberculosis. These
 CC proteins include Rv0287, Rv1037c, Rv2346c, Rv2348c, Rv2653c,
 CC Rv2654c, Rv3020c, Rv3444c, Rv3445c, Rv3890c, Rv3891c, Rv3904c and
 CC Rv3905c. These can be used to produce vaccines against, and in the
 CC diagnosis of, tuberculosis (TB) infection. The present sequence is one of
 CC the proteins of the invention

XX Sequence 95 AA;

Query Match 100.0%; Score 97; DB 4; Length 95;
 Best Local Similarity 100.0%; Pred. No. 7.3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAASAIQGNVTSIH 20
 DB 7 NFAGIEAASAIQGNVTSIH 26

RESULT 15

ABG30968

ID ABG30968 standard; protein; 95 AA.

XX AC ABG30968;

DT 21-OCT-2002 (first entry)

XX Mycobacterium tuberculosis ESAT-6.

KW ESAT-6; mycobacterial infection; tuberculosis; pulmonary tuberculosis;
 KW tuberculosis lymphadenitis; extrapulmonary tuberculosis.

XX Mycobacterium tuberculosis.

XX WO200254072-A2.

XX 11-JUL-2002.

XX 08-JAN-2002; 2002WO-GB000055.

PR 08-JAN-2001; 2001GB-00000432.

PR 08-JAN-2001; 2001US-0259868P.

XX (ISIS-) ISIS INNOVATION LTD.

XX Lalvani A;

DR WPI; 2002-583633/62.

XX Determining the progress of a mycobacterial infection, by direct ex vivo
 PT quantitation of ESAT-6-specific T cells.

XX Disclosure; Page 42; 53pp; English.

XX The invention describes a method of determining the efficacy of treatment
 CC for mycobacterial infection (such as pulmonary tuberculosis, tuberculosis
 CC lymphadenitis and extrapulmonary tuberculosis). The method involves
 CC determining the level of T cells specific for a mycobacterial antigen
 CC that has decreased after the treatment and therefore determining the
 CC efficacy of the treatment. The method is useful for determining the
 CC efficacy of treatment for mycobacterial infection, the mycobacterial
 CC infection is Mycobacterium tuberculosis or M.bovis infection. The
 CC invention also describes a method useful for determining the presence of
 CC a latent infection in a sample from the individual for the presence of
 CC cells specific for a mycobacterial antigen. Also described in a method
 CC for determining the effect of an intervention on a mycobacterial
 CC infection in an individual an a method for treating an individual
 CC infected by a mycobacterium. This is the amino acid sequence of
 CC Mycobacterium tuberculosis ESAT-6, the mycobacterial antigen measured in
 CC the invention

XX Sequence 95 AA;

Query Match 100.0%; Score 97; DB 5; Length 95;
 Best Local Similarity 100.0%; Pred. No. 7.3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAASAIQGNVTSIH 20

DB 7 NFAGIEAASAIQGNVTSIH 26

Search completed: March 10, 2004, 12:05:33
 Job time : 49.5641 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 12:11:07 ; Search time 25.1282 Seconds
(without alignments)
168.061 Million cell updates/sec

Title: US-10-044-703-61

Perfect score: 97
Sequence: 1 NFAGIEAASAIQGNVTSIH 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97	100.0	20	9	US-09-813-333-61
2	97	100.0	20	13	US-10-044-703-61
3	97	100.0	51	14	US-10-193-002-99
4	97	100.0	51	14	US-10-084-843-104
5	97	100.0	51	14	US-10-098-732A-33
6	97	100.0	95	9	US-09-805-427A-1
7	97	100.0	95	11	US-09-872-505-1
8	97	100.0	403	9	US-09-791-171-173
9	97	100.0	403	9	US-09-805-427A-4
10	97	100.0	403	10	US-09-804-980-173
11	97	100.0	403	11	US-09-872-505-4
12	97	100.0	404	9	US-09-791-171-172
13	97	100.0	404	9	US-09-805-427A-3
14	97	100.0	404	10	US-09-804-980-172
15	97	100.0	404	11	US-09-872-505-3

16	85	87.6	20	9	US-09-813-333-60	Sequence 60, Appl
17	85	87.6	20	13	US-10-044-703-60	Sequence 60, Appl
18	66	68.0	15	14	US-10-079-167-80	Sequence 80, Appl
19	66	68.0	15	15	US-10-345-000-2	Sequence 2, Appl
20	60	61.9	13	9	US-09-813-333-59	Sequence 59, Appl
21	60	61.9	13	13	US-10-044-703-59	Sequence 59, Appl
22	47	48.5	443	15	US-10-369-493-9058	Sequence 9058, Ap
23	46	47.4	20	9	US-09-813-333-62	Sequence 62, Appl
24	46	47.4	288	15	US-10-044-703-62	Sequence 62, Appl
25	46	47.4	288	15	US-10-369-493-9706	Sequence 9706, Ap
26	46	47.4	682	9	US-09-798-791-2	Sequence 2, Appl
27	45	46.4	453	15	US-10-369-493-11651	Sequence 11651, A
28	45	46.4	471	15	US-10-369-493-14225	Sequence 14225, A
29	45	46.4	471	15	US-10-369-493-15180	Sequence 15180, A
30	45	46.4	479	15	US-10-369-493-14897	Sequence 14897, A
31	44	45.4	282	15	US-10-369-493-3062	Sequence 3062, Ap
32	44	45.4	319	15	US-10-289-762-893	Sequence 893, App
33	44	45.4	492	15	US-10-369-493-1060	Sequence 1060, Ap
34	43	44.3	15	14	US-10-079-167-79	Sequence 79, Appl
35	43	44.3	15	15	US-10-345-000-1	Sequence 1, Appl
36	43	44.3	239	14	US-10-103-313-351	Sequence 351, App
37	43	44.3	577	14	US-10-317-832-6	Sequence 6, Appl
38	43	44.3	943	16	US-10-389-566-756	Sequence 756, App
39	43	44.3	943	16	US-10-389-566-2231	Sequence 2231, Ap
40	42	43.3	9	9	US-09-916-201-14	Sequence 14, Appl
41	42	43.3	407	14	US-10-156-761-11049	Sequence 11049, A
42	41	42.3	149	15	US-10-264-213-168	Sequence 168, App
43	41	42.3	282	14	US-10-288-930-113	Sequence 113, App
44	41	42.3	417	9	US-09-879-959-2	Sequence 2, Appl
45	41	42.3	417	14	US-10-011-769B-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-813-333-61
; Sequence 61, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:
; APPLICANT: DeGroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-61

Query Match 100.0%; Score 97; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	NFAGIEAASAIQGNVTSIH	20
DB	1	NFAGIEAASAIQGNVTSIH	20

RESULT 2

US-10-044-703-61
; Sequence 61, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:
; APPLICANT: DeGroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703

;; CURRENT FILING DATE: 2002-05-20
;; PRIOR APPLICATION NUMBER: 60/190,834
;; PRIOR FILING DATE: 2000-03-20
;; NUMBER OF SEQ ID NOS: 81
;; SOFTWARE: Patent in Ver. 2.1
;; SEQ ID NO 61
;; LENGTH: 20
;; TYPE: PRT
;; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-61

Query Match 100.0%; Score 97; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAAASAIQGNVTSIH 20
DB 1 NFAGIEAAASAIQGNVTSIH 20

RESULT 3
US-10-193-002-99
; Sequence 99, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS

NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.

REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 99:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 99:

US-10-193-002-99

Query Match 100.0%; Score 97; DB 14; Length 51;

Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NFAGIEAAASAIQGNVTSIH 20
DB 7 NFAGIEAAASAIQGNVTSIH 26

RESULT 4
US-10-084-843-104
; Sequence 104, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.

REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-084-843-104

Query Match 100.0%; Score 97; DB 14; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAAASAIQGNVTSIH 20
DB 7 NFAGIEAAASAIQGNVTSIH 26

RESULT 5
US-10-098-732A-33
; Sequence 33, Application US/10098732A
; Publication No. US20030175294A1

```

; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Gudarian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: ESAT-6
; US-10-098-732A-33

Query Match      100.0%; Score 97; DB 14; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.1e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 NFAGIEAAASAIQGNVTSIH 20
Db      7 NFAGIEAAASAIQGNVTSIH 26

RESULT 6
US-09-805-427A-1
; Sequence 1, Application US/09805427A
; Patent No. US20020176867A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; TITLE OF INVENTION: Hybrids of M. tuberculosis Antigens
; FILE REFERENCE: 670001-2002.5
; CURRENT APPLICATION NUMBER: US/09/805,427A
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-805-427A-1

Query Match      100.0%; Score 97; DB 9; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 NFAGIEAAASAIQGNVTSIH 20
Db      7 NFAGIEAAASAIQGNVTSIH 26

RESULT 7
US-09-872-505-1
; Sequence 1, Application US/09872505
; Publication No. US20040013685A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; TITLE OF INVENTION: Nucleic Acid Fragments Derived From M. Tuberculosis
; FILE REFERENCE: 670001-2002.6
; CURRENT APPLICATION NUMBER: US/09/872,505
; CURRENT FILING DATE: 2001-06-01
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-872-505-1

; ORGANISM: Mycobacterium tuberculosis
; US-09-872-505-1

Query Match      100.0%; Score 97; DB 11; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 NFAGIEAAASAIQGNVTSIH 20
Db      7 NFAGIEAAASAIQGNVTSIH 26

RESULT 8
US-09-791-171-173
; Sequence 173, Application US/09791171
; Patent No. US20020094336A1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: CETHINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Birk
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WELDINGH, Karin
; APPLICANT: FLORIO, Walter
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; FILE REFERENCE: 670001-2002.1
; CURRENT APPLICATION NUMBER: US/09/791,171
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 09/050,739
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 0376/97
; PRIOR FILING DATE: 1997-04-02
; PRIOR APPLICATION NUMBER: 1277/97
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/044,624
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: 60/070,488
; PRIOR FILING DATE: 1998-01-05
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 173
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-791-171-173

Query Match      100.0%; Score 97; DB 9; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 NFAGIEAAASAIQGNVTSIH 20
Db      28 NFAGIEAAASAIQGNVTSIH 47

RESULT 9
US-09-805-427A-4
; Sequence 4, Application US/09805427A
; Patent No. US20020176867A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; TITLE OF INVENTION: Hybrids of M. tuberculosis Antigens
; FILE REFERENCE: 670001-2002.5
; CURRENT APPLICATION NUMBER: US/09/805,427A
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

```


OTHER INFORMATION: Recombinant Fusion protein ESAT-6-Ag85B
US-09-805-427A-4

Query Match 100.0%; Score 97; DB 9; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAAASAIQGNVTSIH 20
DB 28 NFAGIEAAASAIQGNVTSIH 47

RESULT 10
US-09-804-980-173
Sequence 173, Application US/09804980
Publication No. US20030147897A1

GENERAL INFORMATION:
APPLICANT: Statens Serum Institut
APPLICANT: Anderson, Peter
TITLE OF INVENTION: M. Tuberculosis Antigens
FILE REFERENCE: 670001-2002.4
CURRENT APPLICATION NUMBER: US/09/804,980
CURRENT FILING DATE: 2001-03-12
NUMBER OF SEQ ID NOS: 257
SOFTWARE: Patent in version 3.0
SEQ ID NO 173
LENGTH: 403
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis

US-09-804-980-173

Query Match 100.0%; Score 97; DB 10; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAAASAIQGNVTSIH 20
DB 28 NFAGIEAAASAIQGNVTSIH 47

RESULT 11
US-09-872-505-4
Sequence 4, Application US/09872505
Publication No. US20040013685A1

GENERAL INFORMATION:
APPLICANT: Statens Serum Institut
TITLE OF INVENTION: Nucleic Acid Fragments Derived From M. Tuberculosis
FILE REFERENCE: 670001-2002.6
CURRENT APPLICATION NUMBER: US/09/872,505
CURRENT FILING DATE: 2001-06-01
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Patent in version 3.0
SEQ ID NO 4
LENGTH: 403
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Fusion protein ESAT-6-Ag85B

US-09-872-505-4

Query Match 100.0%; Score 97; DB 11; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAAASAIQGNVTSIH 20
DB 28 NFAGIEAAASAIQGNVTSIH 47

RESULT 12
US-09-791-171-172
Sequence 172, Application US/09791171
Patent No. US20020094336A1

GENERAL INFORMATION:

APPLICANT: ANDERSEN, Peter
APPLICANT: NIELSEN, Rikke
APPLICANT: OETTINGER, Thomas
APPLICANT: RASMUSSEN, Peter Birk
APPLICANT: ROSENKRANDS, Ida
APPLICANT: WELDINGH, Karin
APPLICANT: FLORIO, Walter
TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
FILE REFERENCE: 670001-2002.1
CURRENT APPLICATION NUMBER: US/09/791,171
CURRENT FILING DATE: 2001-02-20
PRIOR APPLICATION NUMBER: 09/050,739
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 0376/97
PRIOR FILING DATE: 1997-04-02
PRIOR APPLICATION NUMBER: 1277/97
PRIOR FILING DATE: 1997-11-10
PRIOR APPLICATION NUMBER: 60/044,624
PRIOR FILING DATE: 1997-04-18
PRIOR APPLICATION NUMBER: 60/070,488
PRIOR FILING DATE: 1998-01-05
NUMBER OF SEQ ID NOS: 173
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 172
LENGTH: 404
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis

US-09-791-171-172

Query Match 100.0%; Score 97; DB 9; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAAASAIQGNVTSIH 20
DB 316 NFAGIEAAASAIQGNVTSIH 335

RESULT 13
US-09-805-427A-3
Sequence 3, Application US/09805427A
Patent No. US20020176867A1

GENERAL INFORMATION:
APPLICANT: Statens Serum Institut
TITLE OF INVENTION: Hybrids of M. tuberculosis Antigens
FILE REFERENCE: 670001-2002.5
CURRENT APPLICATION NUMBER: US/09/805,427A
CURRENT FILING DATE: 2001-03-13
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Patent in version 3.0
SEQ ID NO 3
LENGTH: 404
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Fusion protein Ag85B-ESAT-6

US-09-805-427A-3

Query Match 100.0%; Score 97; DB 9; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFAGIEAAASAIQGNVTSIH 20
DB 316 NFAGIEAAASAIQGNVTSIH 335

RESULT 14
US-09-804-980-172
Sequence 172, Application US/09804980
Publication No. US20030147897A1

```
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; APPLICANT: Anderson, Peter
; TITLE OF INVENTION: M. Tuberculosis Antigens
; FILE REFERENCE: 670001-2002.4
; CURRENT APPLICATION NUMBER: US/09/804,980
; CURRENT FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 257
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 172
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-804-980-172

Query Match          100.0%; Score 97; DB 10; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 NFAGIEAASAIQGNVTSIH 20
Db      316 NFAGIEAASAIQGNVTSIH 335
      |||||
      |||||

RESULT 15
US-09-872-505-3
; Sequence 3, Application US/09872505
; Publication No. US20040013685A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; TITLE OF INVENTION: Nucleic Acid Fragments Derived From M. Tuberculosis
; FILE REFERENCE: 670001-2002.6
; CURRENT APPLICATION NUMBER: US/09/872,505
; CURRENT FILING DATE: 2001-06-01
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Fusion protein Ag85B-ESAT-6
; US-09-872-505-3

Query Match          100.0%; Score 97; DB 11; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 NFAGIEAASAIQGNVTSIH 20
Db      316 NFAGIEAASAIQGNVTSIH 335
      |||||
      |||||

Search completed: March 10, 2004, 12:41:31
Job time : 25.1282 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:58:01 ; Search time 10.7692 Seconds

(without alignments)
178.641 Million cell updates/sec

Title: US-10-044-703-61

Perfect score: 97

Sequence: 1 NFAGIEAAASAIQGNVTSIH 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78:**

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	100.0	95	2 A70803	early secretory an
2	55	56.7	455	2 JC7550	glucokinase (EC 2.
3	47	48.5	540	2 F86745	ABC transporter pe
4	45	46.4	225	2 T50707	urease accessory p
5	45	46.4	301	2 A82980	acetylglutamate ki
6	45	46.4	487	2 AH2974	aldehyde dehydroge
7	45	46.4	538	2 C98308	aldehyde dehydroge
8	45	46.4	2154	2 A84669	hypothetical prote
9	44.5	45.9	150	2 E97254	protein-tyrosine-p
10	44	45.4	282	2 C72211	acetylglutamate ki
11	44	45.4	307	2 F86594	lipocate synthetase
12	44	45.4	307	2 B72031	lipic acid synthe
13	44	45.4	211	2 F71500	probable lipocate s
14	44	45.4	467	2 JC7551	glucokinase (EC 2.
15	44	45.4	492	2 F64464	sodium-dependent n
16	44	45.4	592	2 S43570	COSB5.6 protein (C
17	44	45.4	774	2 JC7887	alpha,alpha-trehal
18	43.5	44.8	534	2 A69284	coenzyme F420-quin
19	43	44.3	308	2 C81658	lipic acid synthe
20	43	44.3	844	2 B84773	probable mitochond
21	43	44.3	1562	2 T07323	DNA-directed RNA p
22	42.5	43.8	290	2 T16540	hypothetical prote
23	42	43.3	1335	2 T17508	glycoprotein Vp260
24	42	43.3	3413	2 T17467	rifamycin polyketi
25	41	42.3	231	2 T30792	hypothetical prote
26	41	42.3	260	2 AG1871	uroporphyrin-III C
27	41	42.3	215	2 T47971	seven in absentia-
28	41	42.3	427	2 I47126	nuclear RNA helica
29	41	42.3	429	2 A25145	keratin, 47K type

RESULT 1

A70803

early secretory antigen target esat6 - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C/Accession: A70803; S49174

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:96344230

A/Accession: A70803

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-95 <COL>

A/Cross-references: GB:AL022120; GB:AL123456; NID:93261558; PIDN:CAA17967.1; FID:e126412f

A/Experimental source: strain H37RV

R/Soerensen, A.L.; Nagai, S.; Houen, G.; Andersen, P.; Andersen, A.B.

submitted to the EMBL Data Library, June 1994

A/Reference number: S49174

A/Accession: S49174

A/Molecule type: DNA

A/Residues: 1-13, 'R', 15-22, 'S', 24-95 <SOE>

A/Cross-references: EMBL:X79562; NID:9531708

C/Genetics:

A/Gene: esat6

Query Match 100.0%; Score 97; DB 2; Length 95;

Best Local Similarity 100.0%; Pred.No. 2.7e-08; Mismatches 0; Indels 0; Gaps 0;

Matches 20; Conservative 0;

Qy 1 NFAGIEAAASAIQGNVTSIH 20

|||||

Db 7 NFAGIEAAASAIQGNVTSIH 26

|||||

RESULT 2

JC7550

glucokinase (EC 2.7.1.2) ADP-dependent (AMP-forming) - Pyrococcus furiosus

C/Species: Pyrococcus furiosus

C/Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001

C/Accession: JC7550

R/Koga, S.; Yoshitaka, I.; Sakuraba, H.; Takahashi, M.; Sakasegawa, S.; Shimizu, S.; Ohshij

J. Biochem. 128, 1079-1085, 2000

A/Title: Biochemical characterization, cloning, and sequencing of ADP-dependent (AMP-form

A/Reference number: JC7550; MUID: 21062543; PMID:11098152

A/Accession: JC7550

A/Molecule type: DNA

A/Residues: 1-455 <KOG>

A/Cross-references: GB:E14588

A;Experimental source: strain DSM 3638
 C;Comment: This enzyme, a novel sugar kinase, catalyzes the phosphorylation of glucose u
 ified Embden-Meyerhof pathway.
 C;Genetics:

A;Gene: adh-gk
 C;Superfamily: Pyrococcus horikoshii hypothetical protein PH1645
 C;Keywords: phosphotransferase

Query Match 56.7%; Score 55; DB 2; Length 455;
 Best Local Similarity 55.6%; Pred. No. 0.65;
 Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 FAGIEAAASAIQGNVTSI 19

DB 364 FAALAAARAKNGNIISL 381

RESULT 3

F86745
 ABC transporter permease protein yjJd [imported] - Lactococcus lactis subsp. lactis (str
 C;Species: Lactococcus lactis subsp. lactis
 C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
 C;Accession: F86745

R;Bolotin, A.; Wincker, P.; Mauger, S.; Jalllon, O.; Malarne, K.; Weissenbach, J.; Ehrli
 Genome Res. 11, 731-753, 2001

A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis se
 A;Reference number: A86625; MUID:21235186; PMID:11337471

A;Accession: F86745

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-540 <STO>

A;Cross-references: GB:AB005176; PID:g12723906; PIDN:AAK05064.1; GSPDB:GN00146

A;Experimental source: strain IL403

C;Genetics:

A;Gene: yjJd

Query Match 48.5%; Score 47; DB 2; Length 540;
 Best Local Similarity 71.4%; Pred. No. 14;
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 FAGIEAAASAIQGN 15

DB 339 FVGNETAAASAIRGN 352

RESULT 4

T50707
 urease accessory protein D [imported] - Rhodobacter sphaeroides

C;Species: Rhodobacter sphaeroides

C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 22-Jun-2003

C;Accession: T50707

R;Choudhary, M.; Kaplan, S.

Nucleic Acids Res. 28, 862-867, 2000

A;Title: DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides 2
 A;Reference number: Z25222; MUID:20115911; PMID:10648776

A;Accession: T50707

A;Molecule type: translated from GB/EMBL/DBJ

A;Status: preliminary; DNA

A;Residues: 1-225 <CHO>

A;Cross-references: EMBL:AF195122; PIDN:AAF24251.1

A;Experimental source: strain 2.4.1

C;Genetics:

A;Gene: ureD

C;Superfamily: urease accessory protein UreH/UreD

Query Match 46.4%; Score 45; DB 2; Length 225;
 Best Local Similarity 64.3%; Pred. No. 12;
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 AGIEAAASAIQGNV 16

DB 178 AGVEAAASAFDGL 191

RESULT 5

A82980
 acetylglutamate kinase PA5323 [imported] - Pseudomonas aeruginosa (strain PAO1)

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C;Accession: A82980

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Bri
 adman, S.; Yuan, Y.; Brody, L.D.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam,
 ; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
 A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: A82980

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-301 <STO>

A;Cross-references: GB:AE004945; GB:AE004091; NID:g9951639; PIDN:AAG08708.1; GSPDB:GN001;

A;Experimental source: strain PAO1

C;Genetics:

A;Gene: argB; PA5323

C;Superfamily: acetylglutamate kinase

Query Match 46.4%; Score 45; DB 2; Length 301;

Best Local Similarity 56.2%; Pred. No. 16;

Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 5 IEAAASAIQGNVTSIH 20

DB 256 IRCALNAVQGVTSAH 271

RESULT 6

AH2974

aldehyde dehydrogenase Atu3401 [imported] - Agrobacterium tumefaciens (strain C58, Dupont
 C;Species: Agrobacterium tumefaciens

C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C;Accession: AH2974

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCrellan;

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E

ster, E.W.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A;Reference number: AB2577; MUID:21608550; PMID:11743193

A;Accession: AH2974

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-487 <KUR>

A;Cross-references: GB:AE008689; PIDN:AAL44214.1; PID:g17741794; GSPDB:GN00187

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: Atu3401

A;Map position: linear chromosome

C;Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 46.4%; Score 45; DB 2; Length 487;

Best Local Similarity 62.5%; Pred. No. 27;

Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 AGIEAAASAIQGNVTS 18

DB 272 ADIEKAAAIAGNLS 287

RESULT 7

C98308

aldehyde dehydrogenase dhas [imported] - Agrobacterium tumefaciens (strain C58; Cereon)

C;Species: Agrobacterium tumefaciens

C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002

C;Accession: C98308

R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
 A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens* strain C58
 A;Reference number: A97359; MUID:21608551; PMID:11743194
 A;Accession: C98308
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-538 <KUR>
 A;Cross-references: GB:AE007870; PIDN:AAK89989.1; PID:g15159956; GSPDB:GN00170
 C;Genetics:
 A;Gene: AGR L 2842
 A;Map position: linear chromosome
 C;Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 46.4%; Score 45; DB 2; Length 538;
 Best Local Similarity 62.5%; Pred. No. 30;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 AGIEAASAIQGNVTSI 18
 |||||
 Db 323 ADIEKAAAIAGNLS 338
 |||||

RESULT 8
 A84669
 hypothetical protein At2g27120 [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Feb-2001 #sequence_revision 02-feb-2001 #text_change 16-Feb-2001
 C;Accession: A84669
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A;Reference number: A84420; MUID:20083487; PMID:10617197
 A;Accession: A84669
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-2154 <STO>
 A;Cross-references: GB:AE002093; NID:g3885342; PIDN:AAC77870.1; GSPDB:GN00139
 C;Genetics:
 A;Gene: At2g27120
 A;Map position: 2
 C;Superfamily: DNA-directed DNA polymerase II

Query Match 46.4%; Score 45; DB 2; Length 2154;
 Best Local Similarity 47.4%; Pred. No. 1.3e+02;
 Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 NFAGIEAASAIQGNVTSI 19
 |||||
 Db 1860 NYAGIRATGEISGNEVTI 1878
 |||||

RESULT 9
 E97254
 protein-tyrosine-phosphatase, YWLE B. subtilis ortholog [imported] - Clostridium acetobutylicum
 C;Species: Clostridium acetobutylicum
 C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C;Accession: E97254
 R;Nolling, J.; Bretton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4836, 2001
 A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum* ATCC824
 A;Reference number: A96900; MUID:21359325; PMID:21359325
 A;Accession: E97254
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-150 <KUR>
 A;Cross-references: GB:AE001437; PIDN:AAK80824.1; PID:g15025928; GSPDB:GN00168
 A;Experimental source: Clostridium acetobutylicum ATCC824
 C;Genetics:
 A;Gene: CAC2881

Query Match 45.9%; Score 44.5; DB 2; Length 150;
 Best Local Similarity 52.2%; Pred. No. 9.6;
 Matches 12; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

QY 1 NFAGIEAASAIQGNVTSI 20
 |||||
 Db 26 NIDGIEAFSGASAIHGKISLN 48
 |||||

RESULT 10
 C72211
 acetylglutamate kinase - Thermotoga maritima (strain MSB8)
 C;Species: Thermotoga maritima
 C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C;Accession: C72211
 R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M. Nature 399, 323-329, 1999
 A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing
 A;Reference number: A72200; MUID:99287316; PMID:10360571
 A;Accession: C72211
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-282 <ARN>
 A;Cross-references: GB:AE001816; GB:AE000512; NID:g4982359; PIDN:AAD36847.1; PID:g498236;
 A;Experimental source: strain MSB8
 C;Genetics:
 A;Gene: TMI784
 C;Superfamily: acetylglutamate kinase

Query Match 45.4%; Score 44; DB 2; Length 282;
 Best Local Similarity 43.8%; Pred. No. 22;
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 5 IEAASAIQGNVTSI 20
 |||||
 Db 238 VECASVAVRGVGVAVH 253
 |||||

RESULT 11
 P86594
 lipote synthetase [imported] - Chlamydia pneumoniae (strain J138)
 C;Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C;Accession: P86594
 R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise, Nucleic Acids Res. 28, 2311-2314, 2000
 A;Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
 A;Reference number: A86491; MUID:20330349; PMID:10871362
 A;Accession: P86594
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-307 <STO>
 A;Cross-references: GB:BA000008; NID:g8979206; PIDN:BAA99040.1; GSPDB:GN00142
 A;Experimental source: strain J138
 C;Genetics:
 A;Gene: lipA
 C;Superfamily: lipote synthetase

Query Match 45.4%; Score 44; DB 2; Length 307;
 Best Local Similarity 50.0%; Pred. No. 24;
 Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 AGIEAASAIQGNVTSI 20
 |||||
 Db 148 ATTEVLASDFQGNVSAH 165
 |||||

RESULT 12
 B72031
 lipote acid synthetase CP1038 [imported] - Chlamydia pneumoniae (strains CWL029 and J138)

C:Species: Chlamydothila pneumoniae, Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C:Accession: B72031; C81509
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606; PMID:10192388
 A:Accession: B72031
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-307 <ARN>
 A:Cross-references: GB:AE001363; NID:g4377134; PIDN:AA18969.1; PID:g437714
 A:Experimental source: strain CWL029
 R:Read T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Neison, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
 A:Reference number: A81500; MUID:20150255; PMID:10684935
 A:Accession: C81509
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-307 <REA>
 A:Cross-references: GB:AE002261; GB:AE002161; NID:g7189950; PIDN:AAF38813.1; PID:g718995
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: lipA; CP1038
 C:Superfamily: lipoic acid synthase

Query Match 45.4%; Score 44; DB 2; Length 307;
 Best Local Similarity 50.0%; Pred. No. 24;
 Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 3 AGIEAASAIQGNVTSIH 20
 Db 148 ATTEVLASDFGNSALH 165

RESULT 13
 F71500
 probable lipote synthetase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
 C:Accession: F71500
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Pan, J.; Marathe, R.; Aravind, L.; Mitchell,
 Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra
 A:Reference number: A71570; MUID:9900809; PMID:9784136
 A:Accession: F71500
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-311 <ARN>
 A:Cross-references: GB:AE001326; GB:AE001273; NID:g3328990; PIDN:AAC68160.1; PID:g332899
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: lipA
 C:Superfamily: lipoic acid synthase

Query Match 45.4%; Score 44; DB 2; Length 311;
 Best Local Similarity 44.4%; Pred. No. 25;
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 3 AGIEAASAIQGNVTSIH 20
 Db 143 ATIEVLASDFGNTIALH 160

RESULT 14
 JC7551
 glucokinase (EC 2.7.1.2) ADP-dependent (AMP-forming) - Thermococcus litoralis
 C:Species: Thermococcus litoralis
 C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
 C:Accession: JC7551
 R:Koga, S.; Yoshioaka, I.; Sakuraba, H.; Takahashi, M.; Sakasegawa, S.; Shimizu, S.; Ohsh

J. Biochem. 128, 1079-1085, 2000
 A:Title: Biochemical characterization, cloning, and sequencing of ADP-dependent (AMP-form
 A:Reference number: JC7550; MUID: 21062543; PMID:11098152
 A:Accession: JC7551
 A:Molecule type: DNA
 A:Residues: 1-467 <KOG>
 A:Cross-references: GB:E14589
 A:Experimental source: strain DSM 5473
 C:Comment: This enzyme, a novel sugar enzyme, catalyzes the phosphorylation of glucose w
 ified Embden-Meyerhof pathway.
 C:Genetics:
 A:Gene: adp-gk
 C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1645
 C:Keywords: phosphotransferase

Query Match 45.4%; Score 44; DB 2; Length 467;
 Best Local Similarity 50.0%; Pred. No. 37;
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 FAGIEAASAIQGNVTSI 19
 Db 374 PASLAANAAMKGNLERI 391

RESULT 15
 F64464
 sodium-dependent noradrenaline transporter homolog - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: F64464
 R:Butt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
 Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
 A:Reference number: A64300; MUID:96337999; PMID:8688087
 A:Accession: F64464
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-492 <BUL>
 A:Cross-references: GB:U67572; GB:L77117; NID:gi591958; PIDN:AAB99329.1; PID:gi591959; T
 C:Genetics:
 A:Map position: F0E1264962-1266440
 A:Start codon: GTG
 C:Superfamily: gamma-aminobutyric acid transporter

Query Match 45.4%; Score 44; DB 2; Length 492;
 Best Local Similarity 44.4%; Pred. No. 39;
 Matches 8; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy 2 FAGIEAASAIQGNVTSI 19
 Db 330 FAGISSAVSIVEASVSI 347

Search completed: March 10, 2004, 12:12:29
 Job time : 11.7692 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:51:36 ; Search time 6.15385 Seconds
(without alignments)
169.228 Million cell updates/sec

Title: US-10-044-703-61

Perfect score: 97

Sequence: 1 NFAGIEAASAIQGNVTSIH 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	97	100.0	94	1	ESXA_MYCTU
2	45	46.4	300	1	ARGB_PSEAE
3	44	45.4	282	1	ARGB_THEMA
4	44	45.4	307	1	LIPA_CHLPN
5	44	45.4	311	1	LIPA_CHLTP
6	44	45.4	492	1	YD19_METJA
7	44	45.4	592	1	YK06_CAEEL
8	43	44.3	155	1	RIFK_MOUSE
9	43	44.3	162	1	RIFK_HUMAN
10	43	44.3	308	1	LIPA_CHLMU
11	43	44.3	312	1	LIPA_CHLCV
12	43	44.3	577	1	THA4_HUMAN
13	43	44.3	1562	1	RPOD_CHLVU
14	42	43.3	146	1	YD12_BIFLO
15	42	43.3	176	1	DEF1_NITEU
16	42	43.3	383	1	ALR_EAPCS
17	42	43.3	491	1	K5PE_METAC
18	41	42.3	231	1	VF16_VACCP
19	41	42.3	413	1	ALSI_TAMSI
20	41	42.3	415	1	LEU2_SULTO
21	41	42.3	427	1	HE47_PIG
22	41	42.3	429	1	K1C1_XENLA
23	41	42.3	759	1	ARY2_CALVI
24	41	42.3	801	1	BRD2_HUMAN
25	41	42.3	914	1	GNDS_HUMAN
26	41	42.3	956	1	MTN2_MOUSE
27	40.5	41.8	628	1	GIAL_FUSNN
28	40	41.2	196	1	YJ80_HALNI
29	40	41.2	231	1	VF16_VACCC
30	40	41.2	232	1	YDFE_SCHPO
31	40	41.2	257	1	PAEJ_ECOLI
32	40	41.2	301	1	ARGB_PSEPK
33	40	41.2	417	1	WN1A_HUMAN

34 40 41.2 417 1 WN1A_MOUSE
35 40 41.2 418 1 LE21_METKA
36 40 41.2 419 1 HASA_STRPY
37 40 41.2 422 1 CAMA_PSEPU
38 40 41.2 455 1 PTSE_KLEPN
39 40 41.2 513 1 FLAB_HELPJ
40 40 41.2 513 1 FLAB_HELPJ
41 40 41.2 553 1 MCRA_METKA
42 40 41.2 969 1 ARVC_MOUSE
43 40 41.2 1095 1 PIPA_DROME
44 40 41.2 1121 1 YE06_YEAST
45 39 40.2 102 1 CCK2_SYNY3

ALIGNMENTS

RESULT 1
ESXA_MYCTU
ID ESXA_MYCTU STANDARD; PRT; 94 AA.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 6 kDa early secretory antigenic target (ESAT-6).
GN ESXA OR ESAT6 OR RV3875 OR MT3989 OR MTV027.10 OR MB3905.
OS Mycobacterium tuberculosis, and
OC Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=95204931; PubMed=7897219;
RA Andersen P., Andersen A.B., Sorensen A.L., Nagai S.;
RT "Recall of long-lived immunity to Mycobacterium tuberculosis infection
in mice.";
RL J. Immunol. 154:3359-3372(1995).
RN [2]
SEQUENCE FROM N.A., SEQUENCE OF 1-10, AND CHARACTERIZATION.
RC SPECIES=M.tuberculosis; STRAIN=Erdmann;
RX MEDLINE=95247251; PubMed=7729876;
RA Soerensen A.L., Nagai S., Hosen G., Anderson P., Anderson A.B.;
RT "Purification and characterization of a low-molecular-mass T-cell
antigen secreted by Mycobacterium tuberculosis.";
RL Infect. Immun. 63:1710-1717(1995).
RN [3]
SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
Horneby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
RN [4]
SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
Bisai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains.";

J. Bacteriol. 184:5479-5490(2002).
 (5)
 RL SEQUENCE FROM N.A.
 RP SPECIES=M.tuberculosis;
 RC Singh B., Siddiqui Z., Singh S., Sharma P.;
 RA "ESAT-6 gene of a clinical isolate of Mycobacterium tuberculosis from
 RT India";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 (6)
 RN SEQUENCE OF 1-70 FROM N.A.
 RP SPECIES=M.tuberculosis; STRAIN=H37RV;
 RC MEDLINE=9061212; PubMed=9846755;
 RA Berthet F.-X., Rasmussen P.B., Rosenkrands I., Andersen P.,
 RA Gicquel B.;
 RT "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel
 RL low-molecular-mass culture filtrate protein (CFP-10).";
 RL Microbiology 144:3195-3203(1998).
 (7)
 RN SEQUENCE FROM N.A.
 RP SPECIES=M.bovis;
 RC MEDLINE=96200095; PubMed=8631702;
 RA Mahairas G.G., Sabo P.J., Hickey M.J., Singh D.C., Stover C.K.;
 RT "Molecular analysis of genetic differences between Mycobacterium bovis
 RT BCG and virulent M. bovis";
 RL J. Bacteriol. 178:1274-1282(1996).
 (8)
 RN SEQUENCE FROM N.A.
 RP SPECIES=M.bovis; STRAIN=AF2122/97;
 RC MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 CC -!- FUNCTION: NOT KNOWN. ELICITS HIGH LEVEL OF INF-GAMMA FROM MEMORY
 CC EFFECTOR CELLS DURING THE FIRST PHASE OF A PROTECTIVE IMMUNE
 CC RESPONSE.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the ESAT-6 (esx) family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X79562; CAA56099.1; -;
 DR EMBL; AL022120; CAA17967.1; -;
 DR EMBL; AF420491; AAL16896.1; -;
 DR EMBL; AE007190; AAK48357.1; -;
 DR EMBL; AF004671; AAC83446.1; -;
 DR EMBL; U34848; AAC44033.1; -;
 DR EMBL; BX248347; CAD96091.1; -;
 DR PIR; A70803; A70803.
 DR TIGR; MT3989; -;
 DR Tuberculist; Rv3875; -;
 KW Antigen; Complete proteome.
 FT INIT MET 0
 FT SEQUENCE 94 AA; 9773 MW; 19245B0EC478BC84 CRC64;
 Query Match 100.0%; Score 97; DB 1; Length 94;
 Best Local Similarity 100.0%; Pred. No. 1e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NFAGIEAAASAIQGNVTSIH 20
 DB 6 NFAGIEAAASAIQGNVTSIH 25
 RESULT 2

ARGB PSEAE
 ID _ARGB_PSEAE STANDARD; PRT; 300 AA.
 AC Q9HTN2;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Acetylglutamate kinase (EC 2.7.2.8) (NAG kinase) (AGK) (N-acetyl-L-
 DE glutamate 5-phosphotransferase).
 GN ARGB OR PA5323.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=2043737; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 RN [2]
 RP SEQUENCE OF 1-17, AND CRYSTALLIZATION.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=22032476; PubMed=12037312;
 RA Fernandez-Murga M.L., Ramon-Maiques S., Gil-Ortiz F., Fita I.,
 RA Rubio V.;
 RT "Towards structural understanding of feedback control of arginine
 RT biosynthesis: cloning and expression of the gene for the
 RT arginine-inhibited N-acetyl-L-glutamate kinase from Pseudomonas
 RT aeruginosa, purification and crystallization of the recombinant
 RT enzyme and preliminary X-ray studies.";
 RL Acta Crystallogr. D 58:1045-1047(2002).
 CC -!- CATALYTIC ACTIVITY: ATP + N-acetyl-L-glutamate = ADP + N-acetyl-L-
 CC glutamate 5-phosphate.
 CC -!- PATHWAY: Arginine biosynthesis; second step.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -!- MASS SPECTROMETRY: MW=31711; METHOD=MALDI.
 CC -!- SIMILARITY: Belongs to the acetylglutamate kinase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AE004945; AAG08708.1; -;
 DR PIR; A82980; A82980.
 DR HAMAP; MF 00082; -; 1.
 DR InterPro; IPR001048; Aa_kinase.
 DR InterPro; IPR004662; Acg_kinase.
 DR InterPro; IPR001057; Glu_Skinase.
 DR Pfam; PF00696; aak_kinase; 1.
 DR PRINTS; PR00474; GLU5KINASE.
 DR TIGRFAMs; TIGR00761; argB; 1.
 KW Arginine biosynthesis; Transferase; Kinase; Complete proteome.
 FT INIT MET 0
 FT ACT_SITE 32 32 BY SIMILARITY.
 FT ACT_SITE 254 254 BY SIMILARITY.
 FT SITE 87 87 SUBSTRATE BINDING (BY SIMILARITY).
 FT SITE 89 89 SUBSTRATE BINDING (BY SIMILARITY).
 FT SITE 194 194 SUBSTRATE BINDING (BY SIMILARITY).
 FT SEQUENCE 300 AA; 31718 MW; 9CEBD87E49878EB1 CRC64;
 Query Match 46.4%; Score 45; DB 1; Length 300;
 Best Local Similarity 56.2%; Pred. No. 7.4;

Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 5 IEAASAIQGNVTSIH 20
 DB 255 IRCALEAVQGVTSAH 270

RESULT 3
 ARGV_THEME
 ID ARGV_THEME STANDARD; PRT; 282 AA.
 AC Q9X2A4;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Acetylglutamate kinase (EC 2.7.2.8) (NAG kinase) (AGK) (N-acetyl-L-glutamate 5-phosphotransferase).
 GN ARGV OR TM1784.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NSB8 / DSM 3109 / ATCC 43589;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A., McDonald L., Uitterback T.R., Malek J.A., Linher K.D., Garrett M.M., Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D., Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O., Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RA "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima."
 RT Nature 399:323-329(1999).
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K., Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RA "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CML029 from USA."
 RT Nucleic Acids Res. 28:2311-2314(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RA Hirakawa H., Shirai M., Kuhara S.;
 RT "Genomic sequence comparison of two unrelated isolates of Chlamydia pneumoniae from Japan and U.S."
 RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TW-183;
 RA Geng M.M., Schuhmacher A., Muehlendorfer I., Bensch K.W., Schaefer K.P., Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
 RA "The genome sequence of Chlamydia pneumoniae TW183 and comparison with other Chlamydia strains based on whole genome sequence analysis."
 RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Synthesis of alpha-(+)-lipoic acid. It may be involved in the sulfur insertion chemistry in lipoate biosynthesis (By similarity).
 CC -1- PATHWAY: Lipoate biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the biotin and lipoic acid synthetases family.

Query Match 45.4%; Score 44; DB 1; Length 282;
 Best Local Similarity 43.8%; Pred. No. 10;
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 5 IEAASAIQGNVTSIH 20
 DB 238 VECASVAVGGVAVH 253

RESULT 4
 LIPA_CHLPH
 ID LIPA_CHLPH STANDARD; PRT; 307 AA.
 AC Q92774; Q9J018;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Lipoic acid synthetase (lip-syn) (lipoate synthase).
 GN LIPA OR CPN0832 OR CP1038 OR CPB0861.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CML029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W., Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RA "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
 RT Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., White O., Hickey E.K., Peterson J., Uitterback T., Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
 RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39."
 RT Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K., Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RA "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CML029 from USA."
 RT Nucleic Acids Res. 28:2311-2314(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RA Hirakawa H., Shirai M., Kuhara S.;
 RT "Genomic sequence comparison of two unrelated isolates of Chlamydia pneumoniae from Japan and U.S."
 RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TW-183;
 RA Geng M.M., Schuhmacher A., Muehlendorfer I., Bensch K.W., Schaefer K.P., Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
 RA "The genome sequence of Chlamydia pneumoniae TW183 and comparison with other Chlamydia strains based on whole genome sequence analysis."
 RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Synthesis of alpha-(+)-lipoic acid. It may be involved in the sulfur insertion chemistry in lipoate biosynthesis (By similarity).
 CC -1- PATHWAY: Lipoate biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the biotin and lipoic acid synthetases family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

DR EMBL; AB001663; AAD18969.1; -
DR EMBL; AE002261; AAF38813.1; -
DR EMBL; AP002548; BAA99040.1; -
DR EMBL; AB035942; BAA88650.1; -
DR EMBL; AE017160; AAP98790.1; -
DR PIR; B72031; B72031.
DR PIR; F86594; F86594.
DR TIGR; CP1038; -.
DR HAMAP; MF_00206; -.
DR InterPro; IPR006638; Elp3.
DR InterPro; IPR003698; Lipocate synth.
DR InterPro; IPR007197; Radical SAM.
DR Pfam; PF04055; Radical SAM; 1.
DR SMART; SMO0729; Elp3; 1.
DR TIGRFAMs; TIGR00510; lipA; 1.
KW Iron-sulfur; Complete proteome.
FT METAL 78 78 IRON-SULFUR (POTENTIAL).
FT METAL 82 82 IRON-SULFUR (POTENTIAL).
FT METAL 85 85 IRON-SULFUR (POTENTIAL).
SQ SEQUENCE 307 AA; 33993 MW; F893D50F7BC90062 CRC64;

Query Match 45.4%; Score 44; DB 1; Length 307;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 AGIEAASAIQGNVTSIH 20
DB 148 ATTEVLASDFQGNVSALH 165

RESULT 5
LIPA_CHLTR STANDARD; PRT; 311 AA.
ID LIPA_CHLTR STANDARD; PRT; 311 AA.
AC O84562;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lipic acid synthetase (Lip-syn) (Lipoate synthase).
GN LIPA OR CT558.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UW-3/Cx;
RX MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.
RA "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis."
RT Chlamydia trachomatis.
RL Science 282:754-759(1998).
CC -!- FUNCTION: Synthesis of alpha-(+)-lipoic acid. It may be involved in the sulfur insertion chemistry in lipoate biosynthesis (By similarity).
CC -!- PATHWAY: Lipoate biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the biotin and lipoic acid synthetases family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001326; AAC68160.1; -
CC PIR; F71500; F71500.
CC HAMAP; MF_00206; -.
CC InterPro; IPR006638; Elp3.
CC InterPro; IPR003698; Lipocate synth.

```

```

DR InterPro; IPR007197; Radical SAM.
DR Pfam; PF04055; Radical SAM; 1.
DR TIGRFAMs; TIGR00729; Elp3; 1.
DR SMART; SMO0729; Elp3; 1.
DR TIGR; TIGR00510; lipA; 1.
KW Iron-sulfur; Complete proteome.
FT METAL 73 73 IRON-SULFUR (POTENTIAL).
FT METAL 77 77 IRON-SULFUR (POTENTIAL).
FT METAL 80 80 IRON-SULFUR (POTENTIAL).
SQ SEQUENCE 311 AA; 34660 MW; 0B613421B1F330DB CRC64;

Query Match 45.4%; Score 44; DB 1; Length 311;
Best Local Similarity 44.4%; Pred. No. 11;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 AGIEAASAIQGNVTSIH 20
DB 143 ATTEVLASDFEGNIAALH 160

RESULT 6
YD19_METJA STANDARD; PRT; 492 AA.
ID YD19_METJA STANDARD; PRT; 492 AA.
AC Q58715;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical sodium-dependent transporter MJ1319.
DE Hypothetical sodium-dependent transporter MJ1319.
GN MJ1319.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcales; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=9633799; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D., Sutton G.G., Blake J.A., Fitzgerald L.M., Adams M.D., Reich C.I., Kerlavage A.R., Dougherty B.A., Tomb J.F., Weissman J.M., Glodek A., Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A., Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D., Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C., Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M., Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii."
RT Methanococcus jannaschii.
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: Putative sodium-dependent transporter.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the sodium:neurotransmitter symporter (SNF) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; U67572; AAB999329.1; -
CC PIR; F64464; F64464.
CC TIGR; MJ1319; -.
CC InterPro; IPR000175; Na/ntran_symport.
CC Pfam; PF02029; SNF; 1.
CC PRINTS; PR00176; NANEUSMPORT.
CC PRODOM; PD000448; Na/ntran_symport; 1.
CC PROSITE; PS00810; NA_NEUROTRAN_SYMPT_1; 1.
CC PROSITE; PS00754; NA_NEUROTRAN_SYMPT_2; FALSE NEG.
CC PROSITE; PS0267; NA_NEUROTRAN_SYMPT_3; 1.
KW Hypothetical protein; Transport; Transmembrane; Symport;
KW Complete proteome.
RN TRANSMEM 13 33 POTENTIAL.
FT

```

AC	Q8CFV9	Q35471	Q35472	Q9CQ95	
AD	10-OCT-2003	(Rel. 42)	Created		
DT	10-OCT-2003	(Rel. 42)	Last sequence update		
DE	15-FEB-2004	(Rel. 43)	Last annotation update		
DE	Riboflavin kinase (EC 2.7.1.26)	(ATP:riboflavin 5'-phosphotransferase)			
DN	(Flavokinase) (KOI-4)				
GN	RFK				
OS	Mus musculus (Mouse)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
OX	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J;				
RC	TISSUE=Embryonic stem cells, Kidney, and Small intestine;				
RC	MEDLINE=21085660; PubMed=1217851;				
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Iehii Y.,				
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,				
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,				
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,				
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,				
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,				
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barch G.,				
RA	Blake J., Boffelli D., Bojunga N., Garincci P., de Bonaldo M.F.,				
RA	Brownstein M.B., Sult C., Fleischer C., Fujita M., Gariboldi M.F.,				
RA	Gustincich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,				
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,				
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,				
RA	Hayashizaki Y.,				
RT	"Functional annotation of a full-length mouse cDNA collection."				
RL	Nature 409:685-690(2001).				
RL	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=olfactory epithelium, and Retina;				
RC	MEDLINE=22388257; PubMed=12477932;				
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,				
RA	Klausner R.D., Collins F.S., Wagner L., Shenman C.M., Schuler G.D.,				
RA	Altschul S.F., Zeeberg B., Bustow K.H., Schaefer C.F., Bhat N.K.,				
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.J., Haieh P.,				
RA	Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,				
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,				
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,				
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,				
RA	Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,				
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,				
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA	Faney J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,				
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA	Blakesley R.C., Touchman J.W., Green E.D., Dickson M.C.,				
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,				
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,				
RA	Schwerch A., Schein J.E., Jones S.J.M., Marra M.A.;				
RT	"Generation and initial analysis of more than 15,000 full-length				
RT	human and mouse cDNA sequences."				
RT	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).				
RL	[3]				
RP	SEQUENCE OF 79-112 AND 114-155 FROM N.A.				
RC	STRAIN=129/Sv;				
RC	Abrantes E.F., Silva A.M., Reis L.F.L.;				
RL	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.				
CC	-!- FUNCTION: Catalyzes the phosphorylation of riboflavin (vitamin B2)				
CC	to form flavin-mononucleotide (FMN) (By similarity).				
CC	-!- CATALYTIC ACTIVITY: ATP + riboflavin = ADP + FMN				
CC	-!- COFACTOR: Needs Zn(2+) or Mg(2+) for activity (By similarity).				
CC	-!- PATHWAY: FMN and FAD biosynthesis.				
CC	-!- SUBUNIT: Monomer (By similarity).				
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).				
CC	-----				


```

SQ SEQUENCE 162 AA; 18410 MW; E80042E7E5C38ACD CRG64;

Query Match      44.3%; Score 43; DB 1; Length 162;
Best Local Similarity 50.0%; Pred.No.8.5;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 NFAGIEAASAIQGNV 16
    |||...|||
DB 115 NFDLSLISAIQSDI 130
    AC Q9PJZ2 STANDARD; PRT; 308 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lipic acid synthetase (Lip-syn) (Lipoate synthase).
GN LIPA OR TC0847.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_TaxID=83560;
CX [1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / NiGi;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinm M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RA pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -! FUNCTION: Synthesis of alpha-(+)-lipoic acid. It may be involved
CC in the sulfur insertion chemistry in lipoate biosynthesis (By
CC similarity).
CC -! PATHWAY: Lipoate biosynthesis.
CC -! SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -! SIMILARITY: Belongs to the biotin and lipoic acid synthetases
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AE002351; AAF39645.1; -.
CC FICR; C81658; C81658.
CC TIGR; TC0847; -.
CC HAMAP; MF_00206; -.
CC InterPro; IPR006638; Elp3.
CC InterPro; IPR003698; Lipoate_synth.
CC InterPro; IPR007197; Radical_SAM.
CC Pfam; PF04055; Radical_SAM; 1.
CC SMART; SM00729; Elp3; 1.
CC TIGRFAMS; TIGR00510; lipA; 1.
CC Iron-sulfur; Complete Proteome.
CC METAL 74 74 IRON-SULFUR (POTENTIAL).
CC FT METAL 78 78 IRON-SULFUR (POTENTIAL).
CC FT METAL 81 81 IRON-SULFUR (POTENTIAL).
CC SQ SEQUENCE 308 AA; 34278 MW; 617846CB79C12A92 CRG64;

Query Match      44.3%; Score 43; DB 1; Length 308;
Best Local Similarity 44.4%; Pred.No.16;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 AGIEAASAIQGNVTSIH 20
    |||...|||

```

10-OCT-2003 (Rel. 42, Last sequence update)
15-MAR-2004 (Rel. 43, Last annotation update)
THAP domain protein 4 (PP238) (CGI-36).
THAP4.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
Qin W.X., Zhou X.M., Zhang P.P., Jiang H.Q., Huang Y., Wan D.F.,
Gu J.R.;
"Novel human cDNA clones with function of inhibiting cancer cell
growth.";
Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 58-577 FROM N.A.
TISSEP-Brain, and Eye;
MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein W.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Kana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullen J.S.,
Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grilwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smalley D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[3]
RP SEQUENCE OF 221-577 FROM N.A.
MEDLINE=20272150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Chang L.-Y., Liu C.-S., Lin W.-C.;
"Identification of novel human genes evolutionarily conserved in
Caenorhabditis elegans by comparative proteomics.";
Genome Res. 10:703-713 (2000).
[4]
RP SEQUENCE OF 221-577 FROM N.A.
Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
Ninomiya K., Iwavanagi T.;
"NEO human cDNA sequencing project.";
Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
[5]
CC -!- SIMILARITY: Contains 1 THAP domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF258556; AAG23759.1;
DR EMBL; BC000247; AAH00247.1; ALT_INIT.
DR EMBL; BC009439; AAH09439.1; ALT_INIT.
DR EMBL; AK001216; BAA91560.1; ALT_INIT.
DR EMBL; AF132570; AAD27745.1;
DR Genew; HGNC:23187; THAP4.
InterPro; IPR006612; DUF_DM3.
Pfam; PF05485; THAP; 1.
SMART; SM00692; DM3; 1.
Zinc-finger; DNA-binding.
THAP-TYPE.
FT DOMAIN 1 90
FT ZN_FING 4 59
FT CONFLICT 121 121 S -> G (IN REF. 1).
FT CONFLICT 261 261 L -> P (IN REF. 3 AND 4).
FT CONFLICT 413 413 R -> M (IN REF. 4).
SQ SEQUENCE 577 AA; 62889 MW; 40660A5ACDD0A7G3 CRC64;
Query Match 44.3%; Score 43; DB 1; Length 577;
Best Local Similarity 60.0%; Pred. No. 29;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 3 AGIEAASAIQGNVT 17
DB 197 AGDESATSSIEGVT 211
RESULT 13
RPOD_CHLVU
ID RPOD CHLVU STANDARD; PRT; 1562 AA.
AC P12465;
DT 01-OCT-1989 (Rel. 12, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6).
GN RPOC2.
OS Chlorella vulgaris.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC Chlorellaceae; Chlorella.
OC NCBI_TaxID=3077;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=IAM C-27 / Tamiya;
RX MEDLINE=97303241; PubMed=9159184;
RA Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito S.,
Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,
Inamura A., Yoshinaga K., Sugiura M.;
"Complete nucleotide sequence of the chloroplast genome from the
green alga Chlorella vulgaris: the existence of genes possibly
involved in chloroplast division.";
Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972 (1997).
[2]
RN SEQUENCE OF 1323-1562 FROM N.A.
RP STRAIN=IAM C-27 / Tamiya;
RC Yoshinaga K., Ohta T., Suzuki Y., Sugiura M.;
RT "Chlorella chloroplast DNA sequence containing a gene for the large
subunit of ribulose-1, 5-bisphosphate carboxylase and a part of a
possible gene for the beta' subunit of RNA polymerase.";
Plant Mol. Biol. 10:245-250 (1988).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as
CC substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC -!- {RNA} (N).
CC -!- SUBUNIT: In chloroplasts, the RNA polymerase is composed of four
CC subunits: alpha, beta, beta', and beta".
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AB001684; BAA57971.1;
DR EMBL; M20655; AAB54107.1;
DR PIR; J00011; J00011.
DR PIR; T07323; T07323.

```
DR HSP; Q9KWU6; IQM.
DR InterPro; IPR007066; RNA_pol_Rpbi_3.
DR InterPro; IPR007083; RNA_pol_Rpbi_4.
DR InterPro; IPR007081; RNA_pol_Rpbi_5.
DR Pfam; PF04983; RNA_pol_Rpbi_3; 1.
DR Pfam; PF05000; RNA_pol_Rpbi_4; 1.
DR Pfam; PF04998; RNA_pol_Rpbi_5; 1.
DR Transcription; DNA-directed RNA polymerase; Chloroplast.
SQ SEQUENCE 1562 AA; 178710 MW; 631E52784742B06B CRC64;

Query Match 44.3%; Score 43; DB 1; Length 1562;
Best Local Similarity 47.1%; Pred. No. 78;
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 3 AGIEAASAIQGNVTSI 19
Db 133 SGVEAAQALETGLTSV 149

RESULT 14
YD12 BIFLO STANDARD; PRT; 146 AA.
AC Q8G4R4;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical UPF0168 protein BL1312.
GN BL1312.
OS Bifidobacterium longum.
CC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
CC Bifidobacteriaceae; Bifidobacterium.
CX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=2294977; PubMed=12381787;
RA Schell M.A., Karmantantzou M., Snel B., Villanova D., Berger B.,
RA Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
RA Priddle R.D., Arigoni F.
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
CC -!- SIMILARITY: Belongs to the UPF0168 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AEO14759; AAN25112.1; ALT_INIT.
DR HAMAP; MF_00440; -; 1.
DR InterPro; IPR005144; ATP.
DR InterPro; IPR003796; DUF193.
DR Pfam; PF03477; ATP-cone; 1.
DR TIGRFAMs; TIGR00244; TIGR00244; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 146 AA; 16456 MW; 993F4A34C324D9DE CRC64;

Query Match 43.3%; Score 42; DB 1; Length 146;
Best Local Similarity 64.3%; Pred. No. 11;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 NFAGIEAASAIQIG 14
Db 130 NFAGLEDFQSAIDG 143

RESULT 15
DEF1 NITEU STANDARD; PRT; 176 AA.
ID -DEF1_NITEU
```

```
AC Q82TW4;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Peptide deformylase 1 (EC 3.5.1.98) (PDF 1) (Polypeptide deformylase
DE 1).
GN DEF1 OR NE1755.
OS Nitrosomonas europaea.
CC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
CC Nitrosomonadaceae; Nitrosomonas.
CX NCBI_TaxID=915;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19718 / IFO 14298;
RX MEDLINE=22586410; PubMed=12700255;
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Savavedra-Soto L.A.,
RA Arciero D.M., Hommes N.G., Whitaker M.M., Arp D.J.;
RT "Complete genome sequence of the ammonia-oxidizing bacterium and
RT obligate chemolithoautotroph Nitrosomonas europaea.";
RL J. Bacteriol. 185:2759-2773(2003).
CC -!- FUNCTION: Removes the formyl group from the N-terminal Met of
CC newly synthesized proteins. Requires at least a dipeptide for an
CC efficient rate of reaction. N-terminal L-methionine is a
CC prerequisite for activity but the enzyme has broad specificity at
CC other positions (By similarity).
CC -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +
CC methionyl peptide.
CC -!- COFACTOR: Binds 1 iron(II) ion (By similarity).
CC -!- SIMILARITY: Belongs to the polypeptide deformylase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; BX32182; CAB95666.1; -.
DR HAMAP; MF_00163; -; 1.
DR InterPro; IPR000181; Fmet_deformylase.
DR Pfam; PF01327; Pep_deformylase; 1.
DR PRINTS; PR01576; PDEFORMYLASE.
DR ProDom; PD003844; Pep_deformylase; 1.
DR TIGRFAMs; TIGR00079; Pep_deformyl; 1. Complete proteome.
KW Protein biosynthesis; Hydrolase; iron; Complete proteome.
FT ACT_SITE 142 142 BY SIMILARITY.
FT METAL 99 99 IRON (BY SIMILARITY).
FT METAL 141 141 IRON (BY SIMILARITY).
FT METAL 145 145 IRON (BY SIMILARITY).
SQ SEQUENCE 176 AA; 19421 MW; EA054FF4D3B63537 CRC64;

Query Match 43.3%; Score 42; DB 1; Length 176;
Best Local Similarity 36.8%; Pred. No. 13;
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 2 FAGIEAASAIQGNVTSIH 20
Db 117 YQGVDAAGASIDRTVTGFH 135

Search completed: March 10, 2004, 12:06:32
Job time : 7.15385 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:57:36 ; Search time 31.6667 Seconds
(without alignments)
199.275 Million cell updates/sec

Title: US-10-044-703-61

Perfect score: 97
Sequence: 1 NFAGIEAAASAIQGNVTSIH 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_protein.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	55	56.7	455	Q9V2Z6	Q9V2Z6 pyrococcus
2	49	50.5	195	Q8R6N8	Q8R6N8 thermoanaer
3	48	49.5	3765	Q801V0	Q801V0 brachydanio
4	47	48.5	540	Q9CGX6	Q9CGX6 lactococcus
5	47	48.5	2749	Q8SA93	Q8SA93 zea mays (m
6	46	47.4	682	Q9P853	Q9P853 aspergillus
7	45	46.4	61	Q8G115	Q8G115 brucella su
8	45	46.4	225	Q9RFF6	Q9RFF6 rhodobacter
9	45	46.4	269	Q8BGN9	Q8BGN9 pseudomonas
10	45	46.4	488	Q8ENJ3	Q8ENJ3 oceanobacil
11	45	46.4	538	Q8UAH3	Q8UAH3 agrobacteri
12	45	46.4	2031	Q9HJ37	Q9HJ37 thermoplas
13	45	46.4	2154	Q9ZVC8	Q9ZVC8 arabidopsis
14	44.5	45.9	150	Q97F71	Q97F71 clostridium
15	44	45.4	177	Q86DA7	Q86DA7 caenorhabdi
16	44	45.4	418	Q9WXF6	Q9WXF6 lactobacill

17	44	45.4	558	17	Q8TQ96
18	44	45.4	774	2	Q8L164
19	43.5	44.8	88	12	Q8B7T9
20	43.5	44.8	534	17	Q29966
21	43	44.3	200	16	Q8XWF2
22	43	44.3	272	16	Q83CP0
23	43	44.3	307	2	Q8L1U8
24	43	44.3	341	10	Q9XIV5
25	43	44.3	350	16	Q8G5U7
26	43	44.3	415	12	Q68820
27	43	44.3	452	16	Q8EVF6
28	43	44.3	514	16	Q7VF81
29	43	44.3	604	16	Q8EVH4
30	43	44.3	823	10	Q8VZP7
31	43	44.3	844	10	Q8S7U0
32	43	44.3	943	10	Q9SM67
33	43	44.3	1746	10	Q7XQE6
34	42.5	43.8	258	5	Q7Z138
35	42.5	43.8	294	5	Q7Z137
36	42	43.3	120	16	Q8CL90
37	42	43.3	156	5	Q9NDT7
38	42	43.3	158	16	Q8G4R4
39	42	43.3	176	16	Q82TW4
40	42	43.3	231	12	Q80HX3
41	42	43.3	324	16	Q832Q0
42	42	43.3	334	2	Q7WYE3
43	42	43.3	346	2	Q9Z156
44	42	43.3	346	2	Q84I97
45	42	43.3	407	16	Q82HJ6

ALIGNMENTS

RESULT 1

Q9V2Z6 PRELIMINARY; PRT; 455 AA.
AC Q9V2Z6;
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-JUN-2003 (TREMREL. 24, Last annotation update)
DE Glucokinase (ADP-dependent glucokinase).
GN GLKA OR F0312.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 3638;
RX MEDLINE=9340034; PubMed=10409652;
RA Tuininga J.E., Verhees C.H., van der Oost J., Kengen S.W., Stams A.J.,
de Vos W.M.;
RT "Molecular and biochemical characterization of the ADP-dependent
phosphofructokinase from the hyperthermophilic archaeon pyrococcus
furiosus";
RT J. Biol. Chem. 274:21023-21028(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF127910; AAJ48401.1; -
DR EMBL; AE010156; AAL80436.1; -
DR PIR; JC7550; JC7550.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR007666; ADP_PFK_GK.
DR Pfam; PF04587; ADP_PFK_GK; 1-
KW Kinase; Complete proteome.
SQ SEQUENCE 455 AA; 51266 MW; 8262ACBF19BED113 CRC64;

Query Match 56.7%; Score 55; DB 17; Length 455;

Best Local Similarity 55.6%; Pred. No. 3.1; Mismatches 3; Indels 0; Gaps 0;
Matches 10; Conservative 5;

QY 2 FAGIEAASAIQGNVTSI 19
|||:||||:|
Db 364 FAALAAAKAMKGNITSL 381
|||:||||:|

RESULT 2
Q8R6N8 PRELIMINARY; PRT; 195 AA.
ID Q8R6N8
AC Q8R6N8
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein TIE2762.
GN TIE2762.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=1197336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AB013214; AAM25867.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 195 AA; 22314 MW; B884831420BBCB4F CRC64;

Query Match 50.5%; Score 49; DB 16; Length 195;
Best Local Similarity 52.6%; Pred. No. 11;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 NFAGIEAASAIQGNVTSI 19
|||:||||:|
Db 57 NIEGFKVASTIDGNVTAI 75
|||:||||:|

RESULT 3
Q801V0 PRELIMINARY; PRT; 3765 AA.
ID Q801V0
AC Q801V0
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE SI:ZC220F6.1 (Novel protein similar to human dynein heavy chain (DHC))
DE (Fragment).
GN SI:ZC220F6.1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Babbage A.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL807739; CAD87807.1; -.
DR GO; GO:0030286; C:dynein complex; IEA.
DR GO; GO:0008567; F:dynein ATPase activity; IEA.
DR GO; GO:0003777; F:microtubule motor activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0007038; F:microtubule-based movement; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR004273; Dyneinheavy.
DR Pfam; PF03028; Dyneinheavy; 1.
DR SMART; SM00382; AAA; 4.
FT NON_TER 3765 3765

SQ SEQUENCE 3765 AA; 429463 MW; 333077E6741DBF54 CRC64;
Query Match 49.5%; Score 48; DB 13; Length 3765;
Best Local Similarity 40.0%; Pred. No. 3.8e+02;
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 NFAGIEAASAIQGNVTSIH 20
|||:||||:|
Db 1195 NFERLNALALVRGNLPSLH 1214
|||:||||:|

RESULT 4
Q9CGX6 PRELIMINARY; PRT; 540 AA.
ID Q9CGX6
AC Q9CGX6
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE ABC transporter permease protein.
GN YJJD OR IL0966.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Winkler P., Mauger S., Jallion O., Malarne K.,
RA Welzenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
DR EMBL; AB006331; AAK05064.1; -.
DR F86745; F86745.
KW Complete proteome.
SQ SEQUENCE 540 AA; 60283 MW; B896CF783987560E CRC64;

Query Match 48.5%; Score 47; DB 16; Length 540;
Best Local Similarity 71.4%; Pred. No. 69;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 FAGIEAASAIQGN 15
|||:||||:|
Db 339 FVGNETAASAIRGN 352
|||:||||:|

RESULT 5
Q8SA93 PRELIMINARY; PRT; 2749 AA.
ID Q8SA93
AC Q8SA93
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative polypeptide.
GN Z195D10.9.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. B73;
RA Ramakrishna W., Emberton J., SanMiguel P., Bennezen J.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. B73;
RA Doebley J.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. B73;

RA Llaça V., Linton E.W., Young S., Kovchok S., Messing J.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF46646; AAL7599.1; -
 DR GO; GO:0000785; C:chromatin; IEA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
 DR GO; GO:0003682; F:chromatin binding; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006333; P:chromatin assembly/disassembly; IEA.
 DR GO; GO:0006310; P:DNA recombination; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006278; P:RNA dependent DNA replication; IEA.
 DR InterPro; IPR001969; Asparticase_AS.
 DR InterPro; IPR000953; Chromo.
 DR InterPro; IPR009007; Pept_A_acid.
 DR InterPro; IPR005162; Retrotrans_gag.
 DR InterPro; IPR001584; RvE.
 DR InterPro; IPR000477; RvTee.
 DR Pfam; PF03732; Retrotrans_gag; 1.
 DR Pfam; PF00665; rve; 1.
 DR Pfam; PF00078; rvc; 2.
 DR SMART; SM00298; CHROMO; 1.
 DR PROSITE; PS00141; ASP_PROTEASE; 1.
 DR PROSITE; PS0013; CHROMO_2; 1.
 DR PolyProtein; RNA-directed DNA polymerase; Transferase.
 SQ SEQUENCE 2749 AA; 307006 MW; 0A2D92A14515970A CRC64;

Query Match 48.5%; Score 47; DB 10; Length 2749;
 Best Local Similarity 45.0%; Pred. No. 3.9e+02;
 Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 NFAGIEAASAIQGNVTSIH 20
 DB 313 NFAVMOGNWIMQGNISSIH 332

RESULT 6
 QYFP853 PRELIMINARY; PRT; 682 AA.
 ID QYFP853 AC QYFP853
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE Fructosyltransferase.
 GN SFT.
 CS Aspergillus sydowii.
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=75750;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IAM 2544;
 RX MEDLINE=20575231; PubMed=11133467;
 RA Haver A.G., Wendenburg R.;
 RT "Gene Cloning and Functional Characterization by Heterologous
 Expression of the Fructosyltransferase of Aspergillus sydowii IAM
 2544.";
 RL Appl. Environ. Microbiol. 67:363-370(2001).
 CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
 DR EMBL; AJ289046; CAB89083.1; -
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR001362; Glyco_hydro_32.
 DR Pfam; PF00251; Glyco_hydro_32; 1.
 DR SMART; SM00640; Glyco_32; 1.
 KW Glycoprotein; Glycosylase; Hydrolase; Transferase.
 SQ SEQUENCE 682 AA; 74665 MW; 91E3FE93106F37FD CRC64;

Query Match 47.4%; Score 46; DB 3; Length 682;
 Best Local Similarity 52.9%; Pred. No. 1.3e+02;

Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 4. GIEAASAIQGNVTSIH 20
 DB 288 GVEGAYAPIQFSVSMH 304

RESULT 7

QYFP815 PRELIMINARY; PRT; 61 AA.
 ID QYFP815 AC QYFP815
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN BR0913
 OS Brucella suis.
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1330 / Biovar 1;
 RX MEDLINE=22247741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Krahl M., Shetty J., Malek J., Van Aken S.E.,
 RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
 RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
 RT "The Brucella suis genome reveals fundamental similarities between
 animal and plant pathogens and symbionts.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
 DR EMBL; AE014394; AAN29841.1; -
 DR TIGR; BR0913; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 61 AA; 6498 MW; 8FA93A8A827B2C0B CRC64;

Query Match 46.4%; Score 45; DB 16; Length 61;
 Best Local Similarity 73.3%; Pred. No. 14;
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 FAGIEAASAIQGNV 16
 DB 34 FAGIKAAASARDGLV 48

RESULT 8
 QYFPF6 PRELIMINARY; PRT; 225 AA.
 ID QYFPF6 AC QYFPF6
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Ured.
 OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
 OC Rhodobacteraceae; Rhodobacter.
 OX NCBI_TaxID=1063;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2.4.1;
 RA Choudhary M., Kaplan S.;
 RT "DNA sequence analysis of the photosynthesis region of Rhodobacter
 sphaeroides 2.4.1.";
 RL Nucleic Acids Res. 0:0-0(2000).
 DR EMBL; AF195122; AAF24251.1; -
 DR PIR; T50707; T50707.
 DR GO; GO:0016151; F:nickel ion binding; IEA.
 DR GO; GO:0006807; P:nitrogen metabolism; IEA.
 DR InterPro; IPR002669; Ured.
 DR Pfam; PF01774; Ured; 1.
 SQ SEQUENCE 225 AA; 23417 MW; 59EC44A05802C29F CRC64;

Query Match 46.4%; Score 45; DB 2; Length 225;
 Best Local Similarity 64.3%; Pred. No. 57;
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 AGIEAASAIQGNV 16
 DB 178 AGVEAASAFDGKL 191

RESULT 9
 Q88GW9 Q88GW9 PRELIMINARY; PRT; 269 AA.
 AC Q88GW9, 2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Conserved hypothetical protein.
 GN PP3598.
 OS Pseudomonas putida (strain KT2440).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 CC Pseudomonadaceae; Pseudomonas.
 CX NCBI_TaxID=160488;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=22423060; PubMed=12534463;
 RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
 RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
 RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
 RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
 RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzes A.,
 RA Uterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
 RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Helm S.,
 RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuemmler B.,
 RA Fraser C.M.;
 RT "Complete genome sequence and comparative analysis of the
 RT metabolically versatile Pseudomonas putida KT2440.";
 RL Environ. Microbiol. 4:799-808(2002).
 DR EMBL; AEO16787; AAN69198.1; -
 DR TIGR; PP3598; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 269 AA; 29580 MW; 8EF58995E3BA2D39 CRC64;

Query Match 46.4%; Score 45; DB 16; Length 269;
 Best Local Similarity 61.4%; Pred. No. 69;
 Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 AGIEAASAIQGNVTSIH 20
 DB 201 AGIEAARAEDGIVEAIIH 218

RESULT 10
 Q8ENJ3 Q8ENJ3 PRELIMINARY; PRT; 488 AA.
 AC Q8ENJ3, 2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Carboxy-terminal processing protease (EC 3.4.99.-).
 GN O82490.
 OS Oceanobacillus iheyensis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
 CX NCBI_TaxID=162710;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H7E831 / DSM 14371 / JCM 11309;
 RX MEDLINE=2220767; PubMed=12235376;
 RA Takami H., Takaki Y., Uchiyama I.;
 RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
 RT Ridge and its unexpected adaptive capabilities to extreme
 RT environments.";
 RL Nucleic Acids Res. 30:3927-3935(2002).

DR EMBL; APO04601; BAC14446.1; -
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0000270; P:peptidoglycan metabolism; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR001478; PDZ; 1.
 DR InterPro; IPR001515; Peptidase_S41.
 DR InterPro; IPR004447; Peptidase_S41A.
 DR InterPro; IPR002477; PG_binding.
 DR InterPro; IPR003581; TSFC.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF03572; Peptidase_S41; 1.
 DR Pfam; PF01471; PG_binding_1; 1.
 DR SMART; SMO0228; PDZ; 1.
 DR SMART; SMO0245; TSFC; 1.
 DR TIGRFAMs; TIGR00225; prc; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Hydrolase; Protease; Complete proteome.
 SQ SEQUENCE 488 AA; 53406 MW; D284FC3C8B34B9F7 CRC64;

Query Match 46.4%; Score 45; DB 16; Length 488;
 Best Local Similarity 47.4%; Pred. No. 1.3e+02;
 Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 NFAGIEAASAIQGNVTSI 19
 DB 120 SPEGIGAEVSMVNGNVTIV 138

RESULT 11
 Q8UAH3 Q8UAH3 PRELIMINARY; PRT; 538 AA.
 ID Q8UAH3
 AC Q8UAH3
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Aldehyde dehydrogenase.
 GN ATU3401 OR AGR_L2842.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
 CX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A., Gordon D.,
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Krespan W., Perry M.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Dolan M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 RT C58.";
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Ourliou B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Hummel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
 RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gordon J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58.";
 RL Science 294:2323-2328(2001).
 DR EMBL; AEO09270; AAL44214.1; ALT_INIT.
 DR EMBL; AEO08342; AAK89989.1; -
 DR FIR; AH2974; AH2974.

```

DR PIR; C98308; C98308.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
DR Complete proteome.
SQ SEQUENCE 538 AA; 56726 MW; 53AD881FC9A1A79F CRC64;

Query Match 46.4%; Score 45; DB 16; Length 538;
Best Local Similarity 62.5%; Pred. No. 1.4e+02;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 AGIEAASAIQGNVTS 18
Db 323 ADIEKAAAIAGNLS 338

RESULT 12
Q9HJ37 PRELIMINARY; PRT; 2081 AA.
AC Q9HJ37;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Conserved hypothetical membrane protein.
GN TA1136.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stöcker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum."
RL Nature 407:508-513(2000).
DR EMBL; AL445066; CAC12262.1; -.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 2081 AA; 227495 MW; 6E02AA6470DB2E9D CRC64;

Query Match 46.4%; Score 45; DB 17; Length 2081;
Best Local Similarity 36.8%; Pred. No. 6e+02;
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 1 NFAGIEAASAIQGNVTSI 19
Db 1252 NISGVPAATVLSNITNI 1270

RESULT 13
Q9ZVC8 PRELIMINARY; PRT; 2154 AA.
ID Q9ZVC8
AC Q9ZVC8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative DNA polymerase epsilon catalytic subunit
DE (EC 2.7.7.7).
GN AT2G2720.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Sukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;

```

```

RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Roo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Unayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RA thaliana."
RL Nature 402:761-768 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE
CC + {DNA} (N).
CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC DIFFERENT REACTIONS OF DNA SYNTHESIS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
DR EMBL; AC005623; AAC77870.1; -.
DR PIR; A84669; A84669.
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003889; F:alpha DNA polymerase activity; IEA.
DR GO; GO:0003890; F:beta DNA polymerase activity; IEA.
DR GO; GO:0003891; F:delta DNA polymerase activity; IEA.
DR GO; GO:0019386; F:deoxycytidyl transferase activity, template. .; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003893; F:epsilon DNA polymerase activity; IEA.
DR GO; GO:0015999; F:eta DNA polymerase activity; IEA.
DR GO; GO:0003895; F:gamma DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016000; F:iota DNA polymerase activity; IEA.
DR GO; GO:0016450; F:kappa DNA polymerase activity; IEA.
DR GO; GO:0016449; F:lambda DNA polymerase activity; IEA.
DR GO; GO:0016448; F:mu DNA polymerase activity; IEA.
DR GO; GO:0016451; F:nu DNA polymerase activity; IEA.
DR GO; GO:0000366; F:nucleotide binding; IEA.
DR GO; GO:0019984; F:sigma DNA polymerase activity; IEA.
DR GO; GO:0016452; F:theta DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0003894; F:zeta DNA polymerase activity; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001917; AminoTrans_I.
DR InterPro; IPR006172; DNA_pol_B.
DR InterPro; IPR006134; DNA_pol_B_dom.
DR InterPro; IPR006133; DNA_pol_B_exo.
DR Pfam; PF00136; DNA_pol_B; 1.
DR Pfam; PF03104; DNA_pol_B_exo; 1.
DR SMART; SM00486; POLBc_1.
DR PROSITE; PS00599; AA TRANSFER CLASS 2; 1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase;
KW Transferase.
SQ SEQUENCE 2154 AA; 246918 MW; ED30FCB35842A574 CRC64;

Query Match 46.4%; Score 45; DB 10; Length 2154;
Best Local Similarity 47.4%; Pred. No. 6.2e+02;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 NFAGIEAASAIQGNVTSI 19
Db 1860 NYAGIRATGDEISGNEVTI 1878

RESULT 14
Q97F71 PRELIMINARY; PRT; 150 AA.
AC Q97F71;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Protein-tyrosine-phosphatase, YMLE B.subtilis ortholog.

```

Search completed: March 10, 2004, 12:10:55
Job time : 33.7917 secs

GN CAC2881.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Onelchenko M.V., Makarova K.S., Zeng Q.,
Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007785; AAKS0824.1; -;
DR PIR; E97254;
DR GO; GO:0004725; F:protein tyrosine phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR00108; Low_mwt_PTPase.
DR Pfam; PF01451; LMWPC; 1.
DR PRINTS; PR00719; LMWPTPASE.
DR SMART; SMO0226; LMWPC; 1.
KW Complete proteome.
SQ SEQUENCE 150 AA; 16734 MW; 080C09D4B9F00B8E CRC64;

Query Match 45.9%; Score 44.5; DB 16; Length 150;
Best Local Similarity 52.2%; Pred. No. 44;
Matches 12; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

Qy 1 NFAGIE--AAASAIOGNVTSIH 20
Db 26 NIDGIEAFSAGASAIHGSKTSLN 48

RESULT 15

Q86DA7
ID Q86DA7 PRELIMINARY; PRT; 177 AA.
AC Q86DA7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein C05B5.8.
GN C05B5.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Mortimore B.J.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z32679; CAD90169.1; -;
DR WormPep; C05B5.8; CE33967.
KW Hypothetical protein.
SQ SEQUENCE 177 AA; 19679 MW; 4B5CA7E05AAD3BE0 CRC64;

Query Match 45.4%; Score 44; DB 5; Length 177;
Best Local Similarity 50.0%; Pred. No. 64;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 NFAGIEAASAIQGNVTS 18
Db 106 NPTGIDSEIEILGNVTT 123

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:50:51 ; Search time 47.5641 Seconds
(without alignments)
118.807 Million cell updates/sec

Title: US-10-044-703-69

Perfect score: 98

Sequence: 1 VSDLKSTAVIPGYPVAGQV 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003ss:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	100.0	20	AAE12284	Mycobacte
2	98	100.0	143	AAW18133	Mycobacte
3	98	100.0	159	AAW63034	Mycobacte
4	98	100.0	159	AAW14325	Mycobacte
5	98	100.0	159	ADA26952	M. tuberc
6	98	100.0	159	AAE39351	M. tuberc
7	68	69.4	20	AAE12283	Mycobacte
8	65	66.3	46	AAW75574	M. tuberc
9	64	65.3	46	AAW75573	M. tuberc
10	64	65.3	47	AAW85674	M. tuberc
11	64	65.3	47	AAW85674	16 KD M.t
12	63	64.3	47	AAW18136	N-termina
13	63	64.3	47	AAW56338	M. tuberc
14	63	64.3	47	AAE393281	M. tuberc
15	62	63.3	47	AAW18185	N-termina
16	62	63.3	47	AAW56337	M. tuberc
17	62	63.3	47	AAE39280	M. tuberc
18	54.5	55.6	47	AAE92887	Mycobacte
19	54.5	55.6	47	ABG74428	M. tuberc
20	46.5	47.4	545	ABBE4094	Drosophil
21	46.5	47.4	545	ABBE4099	Drosophil
22	46.5	47.4	545	AAE03543	Drosophil
23	46.5	47.4	545	ABR82387	D. melano
24	45	45.9	94	AAU52107	Novel hum
25	45	45.9	482	ABG22600	Novel hum

26	45	45.9	688	6	ABB80130	Abb80130 E. coli b
27	44	44.9	136	2	AAW41731	AAW41731 Actinobac
28	44	44.9	139	4	ABB41888	Abb41888 Peptide #
29	44	44.9	139	4	AAW35687	AAW35687 Peptide #
30	44	44.9	139	4	AAW75582	AAW75582 Human bon
31	44	44.9	139	4	AAW62784	AAW62784 Human bra
32	44	44.9	139	4	ABG57324	ABG57324 Human liv
33	44	44.9	139	5	ABG45112	ABG45112 Human pep
34	44	44.9	242	3	AAW44133	AAW44133 Arabidops
35	44	44.9	263	3	AAW44132	AAW44132 Arabidops
36	44	44.9	263	5	AAU93184	AAU93184 Arabidops
37	44	44.9	263	7	ADD30951	ADD30951 Plant yie
38	44	44.9	263	7	ADE37149	ADE37149 Plant yie
39	44	44.9	298	2	AAW86809	AAW86809 Syndecan
40	44	44.9	326	3	AAW44131	AAW44131 Arabidops
41	44	44.9	328	7	ADD26784	ADD26784 Human adi
42	44	44.9	346	4	ABB62884	ABB62884 Drosophil
43	44	44.9	357	4	AAW96237	AAW96237 Putative
44	44	44.9	445	2	AAW27603	AAW27603 Rat neuro
45	44	44.9	445	2	AAW27602	AAW27602 Rat neuro

ALIGNMENTS

RESULT 1

AAE12284

ID AAE12284 standard; peptide; 20 AA.

XX AC AAE12284;

DT 18-DEC-2001 (first entry)

DE Mycobacterium tuberculosis (Mtb) peptide #69.

KW Mycobacterium tuberculosis; Mtb peptide; antibacterial; vaccine;

KW infection; anti-Mtb immune response.

OS Mycobacterium tuberculosis.

PN WO200170774-A2.

PD 27-SEP-2001.

PF 20-MAR-2001; 2001WO-US008906.

PR 20-MAR-2000; 2000US-0190834P.

PA (UYER-) UNIV BROWN RES FOUND.

PI Degroot AS;

DR WPI; 2001-616401/71.

PT New vaccine for immunizing a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis.

PS Disclosure; Fig 4; 42pp; English.

XX The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti- Mycobacterium tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a mammalian subject preferably human. They are used for immunising a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

SQ Sequence 20 AA;

Query Match 100.0%; Score 98; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.1e-08; Mismatches 0; Gaps 0;
Matches 20; Conservative 0; Indels 0; Indels 0; Gaps 0;

Qy 1 VSDLKSTAVIPGYPVAGQV 20
Db 1 VSDLKSTAVIPGYPVAGQV 20

RESULT 2
AAW18163
ID AAW18163 standard; protein; 143 AA.
XX AC AAW18163;
XX DT 13-AUG-1997 (first entry)
XX DE Mycobacterium tuberculosis extracellular 16KD protein.
XX KW Vaccine; vaccinating agent; M.tuberculosis; pathogen; bacteria; virus;
XX KW fungus; protozoan; HIV.
XX OS Mycobacterium tuberculosis.
XX PH Key Location/Qualifiers
FT Protein 15..143
FT /label= Mature
XX WO9637219-A1.
XX 28-NOV-1996.
XX 23-MAY-1996; 96WO-US007781.
XX 23-MAY-1995; 95US-00447398.
XX 20-OCT-1995; 95US-00545926.
XX 31-OCT-1995; 95US-00551149.
XX 06-DEC-1995; 95US-00568357.
XX (REGC) UNIV CALIFORNIA.
XX Horwitz MA, Harth G;
XX WPI; 1997-020936/02.
XX N-PSDB; AAT71597.
XX Vaccines derived from M.tuberculosis major abundant extracellular
FT proteins - are easy to prepare and less toxic than conventional killed or
FT attenuated vaccines, useful for protecting against or treating
FT Mycobacterial infections.
XX Claim 33; Page 38; 193pp; English.
XX A vaccinating agent for promoting an immune response in a mammal against
CC Mycobacterium pathogens comprises at least one majorly abundant
CC extracellular protein, i.e. the M.tuberculosis 110, 80, 71, 58, 45, 32A,
CC 32B, 30, 24, 23.5, 23, 16, 14 or 12 kD proteins, or their analogues,
CC homologues and subunits. The present sequence represents the 16 kD
CC protein. The vaccinating agents are used to protect against (or to treat
CC existing) infections by Mycobacterium (especially M. tuberculosis) while
CC the epitopes can also be used to detect presence of an immune response to
CC a Mycobacterium pathogen. The vectors, containing the DNA for the
CC extracellular proteins, are used to transform cells for production of
CC recombinant DNA molecules. More generally the DNA from other pathogens
CC can be used in vaccines, e.g. against other bacteria, viruses, fungi and
CC protozoa. Since different combinations of DNA can be used, a wide range
CC of effective compositions can be produced. They generate a response
CC against the antigens most often found on infected cells during the
CC infection, regardless of the strength or specificity of the immune
CC response. The vaccines are easy to produce and less toxic than known
CC killed or attenuated vaccines, so can be given to immunocompromised
CC subjects, e.g. those with HIV infection
XX Sequence 143 AA;

Query Match 100.0%; Score 98; DB 2; Length 143;
Best Local Similarity 100.0%; Pred. No. 9.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSDLKSTAVIPGYPVAGQV 20
Db 41 VSDLKSTAVIPGYPVAGQV 60

RESULT 3
AAW63034
ID AAW63034 standard; protein; 159 AA.
XX AC AAW63034;
XX DT 23-OCT-1998 (first entry)
XX DE Mycobacterium tuberculosis 16 kD protein sequence.
XX KW Mycobacterium tuberculosis; vaccination; extracellular product;
XX KW immunodominant epitope; interleukin-12; MF59; immune response;
XX KW opsonising humoral response; intracellular pathogen.
XX OS Mycobacterium tuberculosis.
XX WO9831388-A1.
XX 23-JUL-1998.
XX 15-JAN-1998; 98WO-US000942.
XX 21-JAN-1997; 97US-00786533.
XX (REGC) UNIV CALIFORNIA.
XX Horwitz MA, Harth G, Lee B;
XX WPI; 1998-413815/35.
XX N-PSDB; AAV42597.
XX Vaccines against Mycobacterium containing major extracellular proteins -
FT used to, e.g. induce protective and therapeutic immune responses, and for
FT detecting an immune response.
XX Example 2; Page 40-41; 236pp; English.
XX This represents a Mycobacterium tuberculosis 16 kD protein. The invention
CC provides an agent for vaccinating mammals against Mycobacterium. The
CC agent comprises at least one of the major abundant extracellular 110, 80,
CC 71, 58, 45, 32A, 32B, 30, 24, 23.5, 23, 16, 14 or 12 kD proteins of M.
CC tuberculosis or at least 1 of their immunodominant epitopes and
CC interleukin-12 (IL-12) or MF59 as adjuvants. The agent containing the
CC nucleic acid encoding the extracellular products are used to raise a
CC protective or therapeutic immune response against Mycobacterium,
CC specifically M. tuberculosis. The immunodominant epitopes can also be
CC used (typically in a cutaneous hypersensitivity test) to detect an immune
CC response to vaccination. Preparation of the agent does not require
CC selection of the most immunogenic products, so large scale production and
CC purification are easy, resulting in a consistent, standardised
CC formulation, having lower toxicity than killed or attenuated vaccines.
CC The agents provide a rapid and effective response (including a strong
CC cell-mediated component) and are safe even in immunocompromised subjects.
CC They prevent development of an opsonising humoral response that might
CC spread intracellular pathogens
XX Sequence 159 AA;

Query Match 100.0%; Score 98; DB 2; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSDLKSTAVIPGYPVAGQV 20


```

DE XX M. tuberculosis extracellular 16 kD protein.
KW XX Vaccine; antibacterial, fungicide; protozoacide; immunostimulant;
KW XX virucide; therapy.
OS XX
XX XX Mycobacterium tuberculosis.
XX XX
XX XX US6599510-B1.
XX XX
XX XX 29-JUL-2003.
XX XX
XX XX 21-SEP-1998; 98US-00157689.
XX XX
XX XX 23-NOV-1993; 93US-00156358.
XX XX 12-AUG-1994; 94US-00289667.
XX XX 23-MAY-1995; 95US-00447398.
XX XX 31-OCT-1995; 95US-00551149.
XX XX 06-DEC-1995; 95US-00568357.
XX XX 23-MAY-1996; 96US-00652842.
XX XX
XX XX (REGC ) UNIV CALIFORNIA.
XX XX
XX XX Horwitz MA, Harth G;
XX XX
XX XX WPI; 2003-669607/63.
XX XX
XX XX N-PSDB; AAD59702.
XX XX
XX XX A nucleic acid encoding an abundant extracellular protein of
XX XX Mycobacterium tuberculosis useful as vaccines for generating protective
XX XX or therapeutic immune response against viral, bacterial, fungal or
XX XX protozoal infections.
XX XX
XX XX Example 2; Col 29-30; 82pp; English.
XX XX
XX XX The invention relates to a novel nucleic acid encoding an abundant
XX XX extracellular protein of Mycobacterium tuberculosis useful as vaccines
XX XX for generating protective or therapeutic immune response against viral,
XX XX bacterial, fungal and protozoal infections. They are also used as
XX XX immunotherapeutic agents. The present sequence is M. tuberculosis strain
XX XX Erdman extracellular protein
XX XX
XX XX Sequence 159 AA;
XX XX
XX XX Query Match 100.0%; Score 98; DB 7; Length 159;
XX XX Best Local Similarity 100.0%; Pred. NO. 1.1e-07;
XX XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX XX
XX XX 1 VSDLKSTAVIPGYPVAGQV 20
XX XX |||||
XX XX 56 VSDLKSTAVIPGYPVAGQV 75
XX XX
XX XX RESULT 7
XX XX AAE12283
XX XX ID AAE12283 standard; peptide; 20 AA.
XX XX
XX XX AC AAE12283;
XX XX
XX XX 18-DEC-2001 (first entry)
XX XX
XX XX Mycobacterium tuberculosis (Mtb) peptide #68.
XX XX
XX XX Mycobacterium tuberculosis; Mtb peptide; antibacterial; vaccine;
XX XX infection; anti-Mtb immune response.
XX XX
XX XX Mycobacterium tuberculosis.
XX XX
XX XX WO200170774-A2.
XX XX
XX XX 27-SEP-2001.
XX XX
XX XX 20-MAR-2001; 2001WO-US008906.
XX XX

```

```

PR XX 20-MAR-2000; 2000US-0190834P.
XX XX (UYBR-) UNIV BROWN RES FOUND.
XX XX
XX XX Degroot AS;
XX XX
XX XX WPI; 2001-616401/71.
XX XX
XX XX New vaccine for immunizing a mammalian subject, preferably humans,
XX XX against infection caused by Mycobacterium tuberculosis.
XX XX
XX XX Disclosure; Fig 4; 42pp; English.
XX XX
XX XX The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine
XX XX candidate peptides. The invention also relates to a method for
XX XX identifying Mtb vaccine candidate peptides as well as vaccines comprising
XX XX these candidate peptides. Vaccines of the invention and Mtb vaccine
XX XX candidate peptides are useful for inducing an anti- Mycobacterium
XX XX tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a
XX XX mammalian subject preferably human. They are used for immunising a
XX XX mammalian subject, preferably human, against infection caused by
XX XX Mycobacterium tuberculosis. The present sequence is a Mtb vaccine
XX XX candidate peptide
XX XX
XX XX Sequence 20 AA;
XX XX
XX XX Query Match 69.4%; Score 68; DB 4; Length 20;
XX XX Best Local Similarity 100.0%; Pred. NO. 0.0008;
XX XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX XX
XX XX 1 VSDLKSTAVIPGY 14
XX XX |||||
XX XX 7 VSDLKSTAVIPGY 20
XX XX
XX XX RESULT 8
XX XX AAW75574
XX XX ID AAW75574 standard; protein; 46 AA.
XX XX
XX XX AC AAW75574;
XX XX
XX XX 23-OCT-1998 (first entry)
XX XX
XX XX M. tuberculosis 16 kD extracellular product N-terminal sequence 2.
XX XX
XX XX Mycobacterium tuberculosis; vaccination; extracellular product;
XX XX immunodominant epitope; interleukin-12; M59; immune response;
XX XX opsonising humoral response; intracellular pathogen.
XX XX
XX XX Mycobacterium tuberculosis.
XX XX
XX XX WO9831388-A1.
XX XX
XX XX 23-JUL-1998.
XX XX
XX XX 15-JAN-1998; 98WO-US000942.
XX XX
XX XX 21-JAN-1997; 97US-00786533.
XX XX
XX XX (REGC ) UNIV CALIFORNIA.
XX XX
XX XX Horwitz MA, Harth G, Lee B;
XX XX
XX XX WPI; 1998-413815/35.
XX XX
XX XX Vaccines against Mycobacterium containing major extracellular proteins -
XX XX used to, e.g. induce protective and therapeutic immune responses, and for
XX XX detecting an immune response.
XX XX
XX XX Example 2; Page 35; 236pp; English.
XX XX
XX XX Sequences shown in AAW75570 to AAW75586 represent N-terminal amino acid
XX XX sequences of 14 exemplary major abundant extracellular products of

```

CC Mycobacterium tuberculosis. The invention provides an agent for
 CC vaccinating mammals against Mycobacterium. The agent comprises at least
 CC one of the major abundant extracellular 110, 80, 71, 58, 45, 32A, 32B,
 CC 30, 24, 23.5, 23, 16, 14 or 12 kDa proteins of M. tuberculosis, or at
 CC least 1 of their immunodominant epitopes and interleukin-12 (IL-12) or
 CC MF59 as adjuvants. The agent containing the nucleic acid encoding the
 CC extracellular products are used to raise a protective or therapeutic
 CC immune response against Mycobacterium, specifically M. tuberculosis. The
 CC immunodominant epitopes can also be used (typically in a cutaneous
 CC hypersensitivity test) to detect an immune response to vaccination.
 CC Preparation of the agent does not require selection of the most
 CC immunogenic products, so large scale production and purification are
 CC easy, resulting in a consistent, standardised formulation, having lower
 CC toxicity than killed or attenuated vaccines. The agents provide a rapid
 CC and effective response (including a strong cell-mediated component) and
 CC are safe even in immunocompromised subjects. They prevent development of
 CC an opsonising humoral response that might spread intracellular pathogens
 CC

XX Sequence 46 AA;
 SQ

Query Match 56.3%; Score 65; DB 2; Length 46;
 Best Local Similarity 70.0%; Pred. No. 0.006;
 Matches 14; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 VSDLKSTAVIPGYPVAGQV 20
 Db 27 VSDLKSTAVIPGYPVAGQV 46

RESULT 9
 AAW75573
 ID AAW75573 standard; protein; 46 AA.

XX AC AAW75573;
 XX DT 23-OCT-1998 (first entry)
 XX DE M. tuberculosis 16 kD extracellular product N-terminal sequence 1.
 XX KW Mycobacterium tuberculosis; vaccination; extracellular product;
 XX KW immunodominant epitope; interleukin-12; MF59; immune response;
 XX KW opsonising humoral response; intracellular pathogen.

XX OS Mycobacterium tuberculosis.
 XX PN WO9831388-A1.
 XX PD 23-JUL-1998.
 XX PF 15-JAN-1998; 98WO-US000942.
 XX PR 21-JAN-1997; 97US-00786533.
 XX PA (REGC) UNIV CALIFORNIA.

XX PI Horwitz MA, Harth G, Lee B;
 XX DR WPI; 1998-413815/35.
 XX PT Vaccines against Mycobacterium containing major extracellular proteins -
 XX PT used to, e.g. induce protective and therapeutic immune responses, and for
 XX PT detecting an immune response.

PS Example 2; Page 35; 236pp; English.

XX CC Sequences shown in AAW75570 to AAW75586 represent N-terminal amino acid
 CC sequences of 14 exemplary major abundant extracellular products of
 CC Mycobacterium tuberculosis. The invention provides an agent for
 CC vaccinating mammals against Mycobacterium. The agent comprises at least
 CC one of the major abundant extracellular 110, 80, 71, 58, 45, 32A, 32B,
 CC 30, 24, 23.5, 23, 16, 14 or 12 kDa proteins of M. tuberculosis, or at
 CC least 1 of their immunodominant epitopes and interleukin-12 (IL-12) or
 CC MF59 as adjuvants. The agent containing the nucleic acid encoding the

CC extracellular products are used to raise a protective or therapeutic
 CC immune response against Mycobacterium, specifically M. tuberculosis. The
 CC immunodominant epitopes can also be used (typically in a cutaneous
 CC hypersensitivity test) to detect an immune response to vaccination.
 CC Preparation of the agent does not require selection of the most
 CC immunogenic products, so large scale production and purification are
 CC easy, resulting in a consistent, standardised formulation, having lower
 CC toxicity than killed or attenuated vaccines. The agents provide a rapid
 CC and effective response (including a strong cell-mediated component) and
 CC are safe even in immunocompromised subjects. They prevent development of
 CC an opsonising humoral response that might spread intracellular pathogens
 CC

XX Sequence 46 AA;
 SQ

Query Match 65.3%; Score 64; DB 2; Length 46;
 Best Local Similarity 70.0%; Pred. No. 0.0087; 5; Indels 0; Gaps 0;
 Matches 14; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 VSDLKSTAVIPGYPVAGQV 20
 Db 27 VSDLKSTAVIPGYPVAGQV 46

RESULT 10
 AAR85674
 ID AAR85674 standard; peptide; 47 AA.

XX AC AAR85674;
 XX DT 25-JAN-1996 (first entry)
 XX DE 16 kD M. tuberculosis extracellular protein N-terminal fragment.
 XX KW Majorly abundant extracellular product; MAP; M. tuberculosis; immunise;
 XX KW vaccinating agent; vaccine; M. bovis; M. marinum; M. kansasii;
 XX KW M. avium intracellulare; M. fortuitum; M. chelonae; M. scrofulaceum; M. leprae;
 XX KW M. africanum; M. ulcerans; M. microti; human; cat; dog; cattle; sheep;
 XX KW horse; pig.

XX OS Mycobacterium tuberculosis.

XX FH Key Location/Qualifiers
 XX FT Misc-difference 28 /note= "Uncertain residue"
 XX FT Misc-difference 31 /note= "Uncertain residue"
 XX FT Misc-difference 41 /label= Phe, Tyr
 XX FT Misc-difference 43 /note= "Uncertain residue"
 XX FT Misc-difference 43 /note= "Unknown residue"

XX PN WO9514713-A2.
 XX PD 01-JUN-1995.
 XX PF 18-NOV-1994; 94WO-US013145.
 XX PR 23-NOV-1993; 93US-00156358.
 XX PR 12-AUG-1994; 94US-00289667.
 XX PA (REGC) UNIV CALIFORNIA.

XX PI Horwitz MA;
 XX DR WPI; 1995-206898/27.

XX PT Vaccine composed of major extracellular products of Mycobacterium - used to
 XX PT promote immune response to various Mycobacterium pathogens, such as M.
 XX PT tuberculosis, M. bovis, M. leprae, etc. in humans, cattle, sheep, etc.
 XX PS Claim 21; Page 105; 123pp; English.

XX CC The sequences given in AAR85672-79 represent N-terminal peptides from

CC majorly abundant extracellular products (MAPs) produced by M.
 CC tuberculosis. These proteins constitute approx. 90% of all of the
 CC proteins released by M.tuberculosis in culture. The MAPs can be used to
 CC immunise a host animal, and due to their release from M. tuberculosis
 CC when it infects the host, they are presented to the host immune system at
 CC high frequency. Due to their profuse and continual presentation to the
 CC infected host's immune system, the most prevalent bacterial extracellular
 CC products provoke a vigorous immune response largely irrespective of their
 CC individual molecular immunogenic characteristics. These peptides may be
 CC used in a vaccinating agent to provide immunity against M.tuberculosis,
 CC M.bovis, M.marinum, M.kansasii, M.aviumintracellulare, M.fortuitum,
 CC M.chelonae, M.scrofulaceum, M.leprae, M.africanum, M.ulcerans, and
 CC M.microti in humans, cats, dogs, cattle, sheep, horses and pigs
 XX
 SQ Sequence 47 AA;
 Query Match 65.3%; Score 64; DB 2; Length 47;
 Best Local Similarity 73.7%; Pred. No. 0.0089;
 Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 VSDLKSTAVIPGYVAGQ 19
 Db 27 VSDLKSTAVIPGYTVXEQ 45
 RESULT 11
 AAR85686
 ID AAR85686 standard; peptide; 47 AA.
 AC AAR85686;
 DT 25-JAN-1996 (first entry)
 DE 16 kD M.tuberculosis extracellular protein N-terminal fragment.
 KW Majorly abundant extracellular product; MAP; M.tuberculosis; immunise;
 KW vaccinating agent; vaccine; M.bovis; M.marinum; M.kansasii;
 KW M.aviumintracellulare; M.fortuitum; M.chelonae; M.scrofulaceum; M.leprae;
 KW M.africanum; M.ulcerans; M.microti; human; cat; dog; cattle; sheep;
 KW horse; pig.
 OS Mycobacterium tuberculosis.
 XX
 XX Key Location/Qualifiers
 FH Key Location/Qualifiers
 FT Misc-difference 28 /note= "Uncertain residue"
 FT FT
 FT Misc-difference 31 /label= Phe, Tyr
 FT FT
 FT Misc-difference 41 /note= "Uncertain residue"
 FT FT
 FT Misc-difference 43 /note= "Unknown residue"
 FT FT
 XX WO9514713-A2.
 PN 01-JUN-1995.
 XX 18-NOV-1994; 94WO-US013145.
 XX 23-NOV-1993; 93US-00156358.
 PR 12-AUG-1994; 94US-00289667.
 XX (REGC) UNIV CALIFORNIA.
 XX Horwitz MA;
 PI WPI; 1995-206898/27.
 DR Vaccine composed of major extracellular prods. of Mycobacterium - used to
 PT promote immune response to various Mycobacterium pathogens, such as M.
 ET tuberculosis, M. bovis, M. leprae, etc. in humans, cattle, sheep, etc.
 XX Claim 13; Page 103; 123pp; English.

XX The sequences given in AAR85680-86 represent N-terminal peptides from
 CC majorly abundant extracellular products (MAPs) produced by M.
 CC tuberculosis which were used in the vaccinating agent of the invention.
 CC MAPs constitute approx. 90% of all of the proteins released by
 CC M.tuberculosis in culture. MAPs can be used to immunise a host animal,
 CC and due to their release from M.tuberculosis when it infects the host,
 CC they are presented to the host immune system at high frequency. Due to
 CC their profuse and continual presentation to the infected host's immune
 CC system, the most prevalent bacterial extracellular products provoke a
 CC vigorous immune response largely irrespective of their individual
 CC molecular immunogenic characteristics. The vaccinating agent may be used
 CC to provide immunity against M. tuberculosis, M.bovis, M.marinum,
 CC M.kansasii, M.aviumintracellulare, M.fortuitum, M.chelonae,
 CC M.scrofulaceum, M.leprae, M.africanum, M.ulcerans, and M.microti in
 CC humans, cats, dogs, cattle, sheep, horses and pigs
 XX
 SQ Sequence 47 AA;
 Query Match 65.3%; Score 64; DB 2; Length 47;
 Best Local Similarity 73.7%; Pred. No. 0.0089;
 Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 VSDLKSTAVIPGYVAGQ 19
 Db 27 VSDLKSTAVIPGYTVXEQ 45
 RESULT 12
 AAW18186
 ID AAW18186 standard; peptide; 47 AA.
 AC AAW18186;
 DT 13-AUG-1997 (first entry)
 DE N-terminal sequence from Mycobacterium tuberculosis 16KD protein.
 XX
 XX Vaccine; vaccinating agent; M.tuberculosis; pathogen; bacteria; virus;
 KW fungus; protozoan; HIV.
 OS Mycobacterium tuberculosis.
 XX
 XX Key Location/Qualifiers
 FH Key Location/Qualifiers
 FT Misc-difference 43 /label= Unspecified
 FT FT
 XX WO9637219-A1.
 XX 28-NOV-1996.
 XX 23-MAY-1996; 96WO-US007781.
 XX 23-MAY-1995; 95US-00447398.
 PR 20-OCT-1995; 95US-00545926.
 PR 31-OCT-1995; 95US-00551149.
 PR 06-DEC-1995; 95US-00568357.
 XX (REGC) UNIV CALIFORNIA.
 XX Horwitz MA, Harth G;
 PI WPI; 1997-020936/02.
 DR Vaccines derived from M.tuberculosis major abundant extracellular
 PT proteins - are easy to prepare and less toxic than conventional killed or
 PT attenuated vaccines, useful for protecting against or treating
 PT Mycobacterial infections.
 XX Example 2; Page 109; 193pp; English.
 PS A vaccinating agent for promoting an immune response in a mammal against
 CC Mycobacterium pathogens comprises at least one immunodominant epitope of

CC at least one majorly abundant extracellular protein, i.e. the
 CC M.tuberculosis 110, 80, 71, 58, 45, 32A, 32B, 30, 24, 23.5, 23, 16, 14 or
 CC 12 kD proteins, or their analogues, homologues and subunits. The present
 CC sequence represents the N-terminal amino acid sequence of the 16 kD
 CC protein. The N-terminal amino acid sequences of the major abundant
 CC extracellular products were determined to provide structural data and to
 CC uncover possible relationships between the proteins. The vaccinating
 CC agents are used to protect against (or to treat existing) infections by
 CC Mycobacterium (especially M.tuberculosis) while the epitopes can also be
 CC used to detect presence of an immune response to a Mycobacterium
 CC pathogen. The vectors, containing the DNA for the extracellular proteins,
 CC are used to transform cells for production of recombinant DNA molecules.
 CC More generally the DNA from other pathogens can be used in vaccines, e.g.
 CC against other bacteria, viruses, fungi and protozoa. Since different
 CC combinations of DNA can be used, a wide range of effective compositions
 CC can be produced. They generate a response against the antigens most often
 CC found on infected cells during the infection, regardless of the strength
 CC or specificity of the immune response. The vaccines are easy to produce
 CC and less toxic than known killed or attenuated vaccines, so can be given
 CC to immunocompromised subjects, e.g. those with HIV infection
 CC
 CC Sequence 47 AA;

Query Match 64.3%; Score 63; DB 2; Length 47;
 Best Local Similarity 73.7%; Pred. No. 0.013;
 Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 VSDLKSTAVIPGYPVAGQ 19
 DB 27 VSDLYKSTAVIPGTYVXEQ 45

RESULT 13

ABU56338
 ID ABU56338 standard; peptide; 47 AA.

AC ABU56338;

DT 31-MAR-2003 (first entry)

DE M. tuberculosis extracellular product, N-terminal peptide #22.

XX Antibacterial; tuberculosis; vaccine; gene therapy.

XX Mycobacterium tuberculosis.

XX US2002131975-A1.

XX 19-SEP-2002.

XX 14-SEP-2001; 2001US-00953510.

XX 23-NOV-1993; 93US-00156358.

XX 12-AUG-1994; 94US-00289667.

XX 23-MAY-1995; 95US-00447398.

XX 31-OCT-1995; 95US-00551149.

XX 06-DEC-1995; 95US-00568357.

XX 23-MAY-1996; 96US-00652842.

XX 21-SEP-1998; 98US-00157689.

XX (REGC) UNIV CALIFORNIA.

XX Horwitz MA, Harth G;

XX WPI; 2003-174073/17.

XX New vaccine, useful for promoting an immune response against infectious

XX pathogens of the genus Mycobacterium in a mammalian host.

XX Example 2; Page 13; 82pp; English.

XX The invention describes a vaccine for promoting an immune response, in a

XX mammalian host, against infectious pathogens of the genus Mycobacterium,

CC comprising at least 1 immunodominant epitope of at least one majority
 CC abundant extracellular product comprising Mycobacterium tuberculosis 110,
 CC 80, 71, 58, 45, 32A, 32B, 30, 24, 23.5, 23, 16, 14, 12 kD protein or
 CC their analogues, homologues or subunits. The proteins and polypeptides of
 CC the invention are useful in gene therapy and treatment of diseases caused
 CC by Mycobacterium such as tuberculosis. This sequence represents an N-
 CC terminal peptide from a Mycobacterium tuberculosis extracellular protein
 CC
 CC Sequence 47 AA;

Query Match 64.3%; Score 63; DB 6; Length 47;

Best Local Similarity 73.7%; Pred. No. 0.013;

Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 VSDLKSTAVIPGYPVAGQ 19

DB 27 VSDLYKSTAVIPGTYVXEQ 45

RESULT 14

AAE39281
 ID AAE39281 standard; peptide; 47 AA.

XX AAE39281;

XX 18-DEC-2003 (first entry)

XX M. tuberculosis extracellular protein N-terminal peptide #19.

XX Vaccine; antibacterial; fungicide; protozoacide; immunostimulant;

XX virucide; therapy.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers

XX Misc-difference 6

XX /note= "This sequence is stated to be the same as that

XX shown as SEQ ID NO: 22 in sequence listing of the

XX specification"

XX FT Misc-difference 43

XX /label= Unknown

XX US6599510-B1.

XX 29-JUL-2003.

XX 21-SEP-1998; 98US-00157689.

XX 23-NOV-1993; 93US-00156358.

XX 12-AUG-1994; 94US-00289667.

XX 23-MAY-1995; 95US-00447398.

XX 31-OCT-1995; 95US-00551149.

XX 06-DEC-1995; 95US-00568357.

XX 23-MAY-1996; 96US-00652842.

XX (REGC) UNIV CALIFORNIA.

XX Horwitz MA, Harth G;

XX WPI; 2003-669607/63.

XX A nucleic acid encoding an abundant extracellular protein of

XX Mycobacterium tuberculosis useful as vaccines for generating protective

XX or therapeutic immune response against viral, bacterial, fungal or

XX protozoal infections.

XX Example 2; Col 24; 82pp; English.

XX The invention relates to a novel nucleic acid encoding an abundant

XX extracellular protein of Mycobacterium tuberculosis useful as vaccines

XX for generating protective or therapeutic immune response against viral,

XX bacterial, fungal and protozoal infections. They are also used as

XX immunotherapeutic agents. The present sequence is M. tuberculosis strain

CC Erdman extracellular protein N-terminal peptide

XX Sequence 47 AA;

Query Match 64.3%; Score 63; DB 7; Length 47;

Best Local Similarity 73.7%; Pred. No. 0.013;

Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 VSDLKSTAVIPGYPVAGQ 19

Db 27 VSDLKSTAVIPGYPVAGQ 45

RESULT 15

AAW18185

ID AAW18185 standard; peptide; 47 AA.

XX AC AAW18185;

DT 13-AUG-1997 (first entry)

DE N-terminal sequence from Mycobacterium tuberculosis 16KD protein.

XX Vaccine; vaccinating agent; M.tuberculosis; pathogen; bacteria; virus;

KW fungus; protozoan; HIV.

XX OS Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT Misc-difference 43

FT /label= Unspecified

XX WO9637219-A1.

XX PD 28-NOV-1996.

XX PF 23-MAY-1996; 96WO-US0007781.

XX PR 23-MAY-1995; 95US-00447398.

XX PR 20-OCT-1995; 95US-00545926.

XX PR 31-OCT-1995; 95US-00551149.

XX PR 06-DEC-1995; 95US-00568357.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Horwitz MA, Harth G;

XX DR WPI; 1997-020936/02.

XX Vaccines derived from M.tuberculosis major abundant extracellular

PT proteins - are easy to prepare and less toxic than conventional killed or

PT attenuated vaccines, useful for protecting against or treating

PT Mycobacterial infections.

XX Example 2; Page 108; 193pp; English.

XX A vaccinating agent for promoting an immune response in a mammal against

CC Mycobacterium pathogens comprises at least one immunodominant epitope of

CC at least one majorly abundant extracellular protein, i.e. the

CC M.tuberculosis 110, 80, 71, 58, 45, 32A, 32B, 30, 24, 23.5, 23, 16, 14 or

CC 12 kD proteins, or their analogues, homologues and subunits. The present

CC sequence represents the N-terminal amino acid sequence of the 16 kD

CC protein. The N-terminal amino acid sequences of the major abundant

CC extracellular products were determined to provide structural data and to

CC uncover possible relationships between the proteins. The vaccinating

CC agents are used to protect against (or to treat existing) infections by

CC Mycobacterium (especially M.tuberculosis) while the epitopes can also be

CC used to detect presence of an immune response to a Mycobacterium

CC pathogen. The vectors, containing the DNA for the extracellular proteins,

CC are used to transform cells for production of recombinant DNA molecules.

CC More generally the DNA from other pathogens can be used in vaccines, e.g.

CC against other bacteria, viruses, fungi and protozoa. Since different

CC combinations of DNA can be used, a wide range of effective compositions

CC can be produced. They generate a response against the antigens most often

CC found on infected cells during the infection, regardless of the strength

CC or specificity of the immune response. The vaccines are easy to produce

CC and less toxic than known killed or attenuated vaccines, so can be given

CC to immunocompromised subjects, e.g. those with HIV infection

XX SQ Sequence 47 AA;

Query Match 63.3%; Score 62; DB 2; Length 47;

Best Local Similarity 73.7%; Pred. No. 0.019;

Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 VSDLKSTAVIPGYPVAGQ 19

Db 27 VSDLKSTAVIPGYPVAGQ 45

Search completed: March 10, 2004, 12:05:34

Job time : 48.5641 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 12:11:07 ; Search time 25.1282 Seconds

(without alignments)
168.061 Million cell updates/sec

Title: US-10-044-703-69

Perfect score: 98

Sequence: 1 VSDLKSTAVIPGYPVAGQV 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:**

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PTC_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PTCUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09C_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	100.0	20	9 US-09-813-333-69	Sequence 69, Appl
2	98	100.0	20	13 US-10-044-703-69	Sequence 69, Appl
3	68	69.4	20	9 US-09-813-333-68	Sequence 68, Appl
4	68	69.4	20	13 US-10-044-703-68	Sequence 68, Appl
5	63	64.3	47	9 US-09-953-510-22	Sequence 22, Appl
6	63	64.3	47	11 US-09-953-413-22	Sequence 22, Appl
7	63	64.3	47	14 US-10-147-255-22	Sequence 22, Appl
8	62	63.3	47	9 US-09-953-510-21	Sequence 21, Appl
9	62	63.3	47	11 US-09-953-413-21	Sequence 21, Appl
10	62	63.3	47	14 US-10-147-255-21	Sequence 21, Appl
11	47	48.0	205	15 US-10-369-493-20483	Sequence 20483, A
12	46.5	47.4	545	14 US-10-053-510-16	Sequence 16, Appl
13	46.5	47.4	545	15 US-10-149-165-6	Sequence 6, Appl
14	46.5	47.4	545	15 US-10-348-052-16	Sequence 16, Appl
15	45	45.9	688	14 US-10-243-789-2	Sequence 2, Appl

16 44 44.9 139 9 US-09-864-761-45205 Sequence 45205, A
17 44 44.9 263 10 US-09-934-455-496 Sequence 496, Appl
18 44 44.9 263 15 US-10-225-068-88 Sequence 88, Appl
19 44 44.9 263 15 US-10-374-780A-452 Sequence 452, Appl
20 44 44.9 328 15 US-10-139-794-129 Sequence 129, Appl
21 44 44.9 416 15 US-10-369-493-8813 Sequence 8813, Appl
22 44 44.9 445 14 US-10-027-049-2 Sequence 2, Appl
23 44 44.9 445 14 US-10-027-049-4 Sequence 4, Appl
24 44 44.9 456 9 US-09-962-646-2 Sequence 2, Appl
25 44 44.9 499 9 US-09-771-956-22 Sequence 22, Appl
26 44 44.9 508 9 US-09-771-956-22 Sequence 22, Appl
27 44 44.9 577 9 US-09-815-242-11760 Sequence 11760, A
28 44 44.9 1558 16 US-10-389-566-1236 Sequence 1236, Appl
29 44 44.9 1604 16 US-10-389-566-1236 Sequence 1236, Appl
30 43 43.9 129 15 US-10-051-874-52 Sequence 52, Appl
31 43 43.9 170 15 US-10-369-493-16665 Sequence 16665, A
32 43 43.9 350 15 US-10-320-797-3123 Sequence 3123, Appl
33 43 43.9 509 14 US-10-097-559-43 Sequence 43, Appl
34 43 43.9 721 15 US-10-369-493-2166 Sequence 2166, Appl
35 42 42.9 170 15 US-10-369-493-23216 Sequence 23216, A
36 42 42.9 267 9 US-09-981-353-82 Sequence 82, Appl
37 42 42.9 299 9 US-09-739-254-73 Sequence 73, Appl
38 42 42.9 299 9 US-09-904-615-73 Sequence 73, Appl
39 42 42.9 299 14 US-10-054-988-73 Sequence 73, Appl
40 42 42.9 299 14 US-10-055-098-73 Sequence 73, Appl
41 42 42.9 415 15 US-10-369-493-18354 Sequence 18354, A
42 42 42.9 445 9 US-09-866-572A-73 Sequence 73, Appl
43 42 42.9 445 9 US-09-866-570A-73 Sequence 73, Appl
44 42 42.9 445 14 US-10-166-984-73 Sequence 73, Appl
45 42 42.9 445 15 US-10-166-984-73 Sequence 73, Appl

ALIGNMENTS

RESULT 1

US-09-813-333-69
; Sequence 69, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:
; APPLICANT: DeGroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 69
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-69

Query Match 100.0%; Score 98; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 2,4e-09; Mismatches 0; Gaps 0;
Matches 20; Conservative 0; Indels 0;

QY 1 VSDLKSTAVIPGYPVAGQV 20

DB 1 VSDLKSTAVIPGYPVAGQV 20

RESULT 2

US-10-044-703-69
; Sequence 69, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:
; APPLICANT: DeGroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters.
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703

; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 69
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-69

Query Match 100.0%; Score 98; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.4e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSDLKSTAVIPGYFVAGQV 20
Db 1 VSDLKSTAVIPGYFVAGQV 20

RESULT 3
US-09-813-333-68
; Sequence 68, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:
; APPLICANT: DeGroet, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; CURRENT APPLICATION NUMBER: US/09/813,333
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 68
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-68

Query Match 69.4%; Score 68; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSDLKSTAVIPGY 14
Db 7 VSDLKSTAVIPGY 20

RESULT 4
US-10-044-703-68
; Sequence 68, Application US/10044703
; Publication No. US2002019223A1
; GENERAL INFORMATION:
; APPLICANT: DeGroet, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 68
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-68

Query Match 69.4%; Score 68; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSDLKSTAVIPGY 14
Db 7 VSDLKSTAVIPGY 20

RESULT 5
US-09-953-510-22
; Sequence 22, Application US/09953510
; Patent No. US20020131975A1
; GENERAL INFORMATION:
; APPLICANT: Horwitz, Marcus A.
; TITLE OF INVENTION: Abundant Extracellular
; Products and Methods for Their Production and Use
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kurt A. MacLean
; STREET: 2029 Century Park East, Suite 3800
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90067

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0,
Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/953,510
FILING DATE: 14-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/447,398
FILING DATE: 23-MAY-1995
APPLICATION NUMBER: US 08/289,667
FILING DATE: 12-AUG-1994
APPLICATION NUMBER: US 08/156,358
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: MacLean, Kurt A.
REGISTRATION NUMBER: 31,118
REFERENCE/DOCKET NUMBER: 112-272
TELECOMMUNICATION INFORMATION:
TELEPHONE: (310) 788-5000
TELEFAX: (310) 277-1297

INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Mycobacterium tuberculosis
STRAIN: Erdman
SEQUENCE DESCRIPTION: SEQ ID NO: 22:
US-09-953-510-22

Query Match 64.3%; Score 63; DB 9; Length 47;
Best Local Similarity 73.7%; Pred. No. 0.0041;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VSDLKSTAVIPGYVAGQ 19
Db 27 VSDLKSTAVIPGYTVXEQ 45

RESULT 6
US-09-953-413-22
; Sequence 22, Application US/09953413

```
; Publication No. US20040018209A1
; GENERAL INFORMATION:
; APPLICANT: Horwitz, Marcus A.
; TITLE OF INVENTION: Abundant Extracellular
; Products and Methods for Their Production and Use
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kurt A. MacLean
; STREET: 2029 Century Park East, Suite 3800
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0,
; Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/953,413
; FILING DATE: 14-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,398
; FILING DATE: 23-MAY-1995
; APPLICATION NUMBER: US 08/289,667
; FILING DATE: 12-AUG-1994
; APPLICATION NUMBER: US 08/156,358
; FILING DATE: 23-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: MacLean, Kurt A.
; REGISTRATION NUMBER: 31,118
; REFERENCE/DOCKET NUMBER: 112-272
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (310) 788-5000
; TELEFAX: (310) 277-1297
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 47 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium tuberculosis
; STRAIN: Erdman
; SEQUENCE DESCRIPTION: SEQ ID NO: 22:
US-09-953-413-22
Query Match 64.3%; Score 63; DB 11; Length 47;
Best Local Similarity 73.7%; Pred. No. 0.0041;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VSDLKSTAVIPGYPVAGQ 19
Db 27 VSDLYKSTAVIPGYTVXEQ 45

RESULT 7
US-10-147-255-22
; Sequence 22, Application US/10147255
; Publication No. US20030152584A1
; GENERAL INFORMATION:
; APPLICANT: Horwitz, Marcus A.
; TITLE OF INVENTION: Abundant Extracellular
; Products and Methods for Their Production and Use
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kurt A. MacLean
; STREET: 2029 Century Park East, Suite 3800
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90067
; COMPUTER READABLE FORM:
```

```
; STREET: 2029 Century Park East, Suite 3800
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0,
; Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/147,255
; FILING DATE: 15-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/226,539A
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/447,398
; FILING DATE: 23-MAY-1995
; APPLICATION NUMBER: US 08/289,667
; FILING DATE: 12-AUG-1994
; APPLICATION NUMBER: US 08/156,358
; FILING DATE: 23-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: MacLean, Kurt A.
; REGISTRATION NUMBER: 31,118
; REFERENCE/DOCKET NUMBER: 112-272
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (310) 788-5000
; TELEFAX: (310) 277-1297
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 47 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium tuberculosis
; STRAIN: Erdman
; SEQUENCE DESCRIPTION: SEQ ID NO: 22:
US-10-147-255-22
Query Match 64.3%; Score 63; DB 14; Length 47;
Best Local Similarity 73.7%; Pred. No. 0.0041;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VSDLKSTAVIPGYPVAGQ 19
Db 27 VSDLYKSTAVIPGYTVXEQ 45

RESULT 8
US-09-953-510-21
; Sequence 21, Application US/09953510
; Patent No. US20020131975A1
; GENERAL INFORMATION:
; APPLICANT: Horwitz, Marcus A.
; TITLE OF INVENTION: Abundant Extracellular
; Products and Methods for Their Production and Use
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kurt A. MacLean
; STREET: 2029 Century Park East, Suite 3800
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90067
; COMPUTER READABLE FORM:
```


MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0,
Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/953,510
FILING DATE: 14-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/447,398
FILING DATE: 23-MAY-1995
APPLICATION NUMBER: US 08/289,667
FILING DATE: 12-AUG-1994
APPLICATION NUMBER: US 08/156,358
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: MacLean, Kurt A.
REGISTRATION NUMBER: 31,118
REFERENCE/DOCKET NUMBER: 112-272
TELECOMMUNICATION INFORMATION:
TELEPHONE: (310) 788-5000
TELEFAX: (310) 277-1297
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Mycobacterium tuberculosis
STRAIN: Erdman
SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-09-953-510-21
Query Match 63.3%; Score 62; DB 9; Length 47;
Best Local Similarity 73.7%; Pred. No. 0.006;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 VSDLKSSTAVIPGYVAGQ 19
Db 27 VSDLFKSTAVIPGYTVXEQ 45
RESULT 9
US-09-953-413-21
Sequence 21, Application US/09953413
Publication No. US20040018209A1
GENERAL INFORMATION:
APPLICANT: Horwitz, Marcus A.
TITLE OF INVENTION: Abundant Extracellular
Products and Methods for Their Production and Use
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kurt A. MacLean
STREET: 2029 Century Park East, Suite 3800
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0,
Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/953,413
FILING DATE: 14-Sep-2001

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/447,398
FILING DATE: 23-MAY-1995
APPLICATION NUMBER: US 08/289,667
FILING DATE: 12-AUG-1994
APPLICATION NUMBER: US 08/156,358
FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: MacLean, Kurt A.
REGISTRATION NUMBER: 31,118
REFERENCE/DOCKET NUMBER: 112-272
TELECOMMUNICATION INFORMATION:
TELEPHONE: (310) 788-5000
TELEFAX: (310) 277-1297
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Mycobacterium tuberculosis
STRAIN: Erdman
SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-09-953-413-21
Query Match 63.3%; Score 62; DB 11; Length 47;
Best Local Similarity 73.7%; Pred. No. 0.006;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 VSDLKSSTAVIPGYVAGQ 19
Db 27 VSDLFKSTAVIPGYTVXEQ 45
RESULT 10
US-10-147-255-21
Sequence 21, Application US/10147255
Publication No. US20030152584A1
GENERAL INFORMATION:
APPLICANT: Horwitz, Marcus A.
TITLE OF INVENTION: Abundant Extracellular
Products and Methods for Their Production and Use
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kurt A. MacLean
STREET: 2029 Century Park East, Suite 3800
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0,
Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/147,255
FILING DATE: 15-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/226,539A
FILING DATE: <Unknown>
APPLICATION NUMBER: US 08/447,398
FILING DATE: 23-MAY-1995
APPLICATION NUMBER: US 08/289,667
FILING DATE: 12-AUG-1994

```

; APPLICATION NUMBER: US 08/156,358
; FILING DATE: 23-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: MacLean, Kurt A.
; REGISTRATION NUMBER: 31,118
; REFERENCE/DOCKET NUMBER: 112-272
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (310) 788-5000
; TELEFAX: (310) 277-1297
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 47 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium tuberculosis
; STRAIN: Erdman
; SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-10-147-255-21

```

```

Query Match      63.3%; Score 62; DB 14; Length 47;
Best Local Similarity 73.7%; Pred. No. 0.006;
Matches 14; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 VSDLKSTAVIPGYPVAGQ 19
    ||||| ||||| |||||
DB 27 VSDLKSTAVIPGYPVKEQ 45

```

RESULT 11

```

US-10-369-493-20483
; Sequence 20483, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20483
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
US-10-369-493-20483

```

```

Query Match      48.0%; Score 47; DB 15; Length 205;
Best Local Similarity 52.9%; Pred. No. 9.9;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 2 SLDKSTAVIPGYPVAG 18
    :|||: |||: |||:
DB 31 NDLKASVRAIPDPYKPG 47

```

RESULT 12

```

US-10-053-510-16
; Sequence 16, Application US/10053510
; Publication No. US20030175939A1
; GENERAL INFORMATION:
; APPLICANT: Saba, Julie D.

```

```

; APPLICANT: Fyrist, Henrik
; TITLE OF INVENTION: SPHINGOSINE-1-PHOSPHATE LYASE POLYPEPTIDES,
; TITLE OF INVENTION: POLYNUCLEOTIDES AND MODULATING AGENTS AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 200116.402C2
; CURRENT APPLICATION NUMBER: US/10/053,510
; CURRENT FILING DATE: 2002-01-17
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-053-510-16

```

```

Query Match      47.4%; Score 46.5; DB 14; Length 545;
Best Local Similarity 43.5%; Pred. No. 37;
Matches 10; Conservative 6; Mismatches 4; Indels 3; Gaps 1;

```

```

QY 1 VSDLKSTAVI---PGYPVAGQV 20
    :||: ||| ||| ||| ||| |||
DB 490 IADVRCTAEIMKDPGPVVGKM 512

```

RESULT 13

```

US-10-149-165-6
; Sequence 6, Application US/10149165
; Publication No. US20030217376A1
; GENERAL INFORMATION:
; APPLICANT: Ebens, Allen J.
; APPLICANT: Keegan, Kevin P.
; APPLICANT: Stout, Thomas J.
; TITLE OF INVENTION: INSECTICIDE TARGETS AND METHODS OF USE
; FILE REFERENCE: GNOP-002
; CURRENT APPLICATION NUMBER: US/10/149,165
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: 60/169,610
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: 60/173,228
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/173,349
; PRIOR FILING DATE: 1999-12-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-149-165-6

```

```

Query Match      47.4%; Score 46.5; DB 15; Length 545;
Best Local Similarity 43.5%; Pred. No. 37;
Matches 10; Conservative 6; Mismatches 4; Indels 3; Gaps 1;

```

```

QY 1 VSDLKSTAVI---PGYPVAGQV 20
    :||: ||| ||| ||| ||| |||
DB 490 IADVRCTAEIMKDPGPVVGKM 512

```

RESULT 14

```

US-10-348-052-16
; Sequence 16, Application US/10348052
; Publication No. US20030219782A1
; GENERAL INFORMATION:
; APPLICANT: Saba, Julie D.
; APPLICANT: Fyrist, Henrik
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE MODULATION
; TITLE OF INVENTION: OF SPHINGOLIPID METABOLISM AND/OR SIGNALING
; FILE REFERENCE: 200116.405
; CURRENT APPLICATION NUMBER: US/10/348,052
; CURRENT FILING DATE: 2003-01-17
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0

```

```
; SEQ ID NO 16
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-348-052-16

Query Match          47.4%; Score 46.5; DB 15; Length 545;
Best Local Similarity 43.5%; Pred. No. 37;
Matches 10; Conservative 6; Mismatches 4; Indels 3; Gaps 1;

QY      1 VSDLKSTAVI---PGYPVAGQV 20
Db      490 IADVRSCTAEIMKDPGQPVVGKM 512

RESULT 15
US-10-243-789-2
; Sequence 2, Application US/10243789
; Publication No. US20030124664A1
; GENERAL INFORMATION:
; APPLICANT: Joly, John C.
; FILE OF INVENTION: AMINOPEPTIDASE
; CURRENT APPLICATION NUMBER: US/10/243,789
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/322,350
; PRIOR FILING DATE: 2001-09-13
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-243-789-2

Query Match          45.9%; Score 45; DB 14; Length 688;
Best Local Similarity 60.0%; Pred. No. 86;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      6 SSTAVICGYPVAGQV 20
Db      489 SQTSLPVIYVAGQV 503
```

Search completed: March 10, 2004, 12:41:31
Job time : 25.1282 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:58:01 ; Search time 10.7692 Seconds
(without alignments)
178.641 Million cell updates/sec

Title: US-10-044-703-69

Perfect score: 98
Sequence: 1 VSDLKSTAVIPGYVAGQV 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*

1: Piri: *
2: Piri2: *
3: Piri3: *
4: Piri4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	100.0	159	2 B70635	hypothetical prote
2	50	51.0	314	2 B83929	response regulator
3	46	46.9	312	2 A87449	conserved hypothet
4	46	46.9	1036	1 T01765	endopeptidase La-1
5	45	45.9	688	2 B65005	hypothetical prote
6	45	45.9	779	2 AG1978	hypothetical prote
7	45	45.9	1787	2 AC2009	serine/threonine k
8	44	44.9	349	2 A86329	F14P1.4 protein -
9	44	44.9	357	2 B75128	GTP-binding protei
10	44	44.9	694	2 F90609	exodeoxyribonuclea
11	44	44.9	1604	2 B86287	F9L1.23 protein -
12	43.5	44.4	940	2 AD1374	internalin protein
13	43	43.9	172	2 E97180	adenine phosphorib
14	43	43.9	327	2 H86939	probable phosphorib
15	43	43.9	476	2 AD1456	hypothetical prote
16	43	43.9	509	2 T51806	cytochrome P450 ka
17	43	43.9	571	2 H70623	probable potassium
18	43	43.9	585	2 T48513	hypothetical prote
19	43	43.9	592	2 C35115	hypothetical prote
20	43	43.9	688	2 H91029	probable peptidase
21	43	43.9	688	2 A85874	probable peptidase
22	43	43.9	721	2 T38665	probable homoconi
23	43	43.9	826	2 AD1683	penicillin-binding
24	42	42.9	118	2 F97479	conserved hypothet
25	42	42.9	170	2 B69587	adenine phosphorib
26	42	42.9	256	2 E72574	probable membrane
27	42	42.9	317	2 T10824	auxin-induced prot
28	42	42.9	401	2 AI3568	mannonate dehydrat
29	42	42.9	410	2 C84827	hypothetical prote

30 42 42.9 412 2 H86708 isocitrate dehydro
31 42 42.9 442 2 T10718 anthranilate N-ben
32 42 42.9 445 2 T10717 anthranilate N-ben
33 42 42.9 445 2 T10719 anthranilate N-ben
34 42 42.9 446 2 T10711 anthranilate N-ben
35 42 42.9 520 2 A35883 laccase [EC 1.10.3
36 42 42.9 520 2 B35883 ligninolytic pheno
37 42 42.9 540 2 C70751 probable fad10 pr
38 42 42.9 686 2 A38235 microtubule-associ
39 42 42.9 862 2 S30911 SMP2 protein - yea
40 42 42.9 1106 1 A39299 DNA-directed DNA p
41 41 41.8 142 2 B70830 hypothetical prote
42 41 41.8 187 2 AD0379 adenine phosphorib
43 41 41.8 298 2 JC6537 mda-9 protein - hu
44 41 41.8 318 2 S37225 ribose-phosphate d
45 41 41.8 357 2 B71003 hypothetical prote

ALIGNMENTS

RESULT 1

B70635
Hypothetical protein Rvi926c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70635

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998

A:Authors: Sqaes, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70635

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-159 <COL>

A:Cross-references: GB:Z84498; GB:AL123456; NID:G3261701; PIDN:CAB06500.1; PID:e293528; 1

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rvi926c

Query Match 100.0%; Score 98; DB:2; Length 159;
Best Local Similarity 100.0%; Pred. No. 9.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSDLKSTAVIPGYVAGQV 20
DB 56 VSDLKSTAVIPGYVAGQV 75

RESULT 2

B83929
response regulatory protein (sensory transduction system) BH2234 [imported] - Bacillus hi

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 12-Jun-2003

C:Accession: B83929

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: B83929

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-314 <STO>

A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA005953.1; GSPDB:GN001

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH2234

C:Superfamily: response regulator diguanylate cyclase, Pld type; response regulator hom

Query Match 51.0%; Score 50; DB 2; Length 314;

```

Best Local Similarity 57.9%; Pred. No. 1.8;
Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 VSDLKSTAVIPGYPVAGQ 19
DB 80 VSSLKEKEAVIKGFEVGQ 98

RESULT 3
A87449
conserved hypothetical protein CC1610 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: A87449
R:NIERMAN, W.C.; FELDBLYUM, T.V.; PAULSEN, I.T.; NELSON, K.E.; EISEN, J.; HEIDELBERG, J.
B.; LAUB, M.T.; DEBOY, R.T.; DODSON, R.J.; DURKIN, A.S.; GWINN, M.L.; HAFT, D.H.; KOLON
N. J.; ERMOLAIEVA, M.; WHITE, O.; SALZBERG, S.L.; SHAPIRO, L.; VENTER, J.C.; FRASER, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4138-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: A87449
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-312 <S>T>
A:Cross-references: GB:AB005673; NID:gl3423009; PIDN:AAK23589.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC1610

Query Match 46.9%; Score 46; DB 2; Length 312;
Best Local Similarity 52.6%; Pred. No. 8.4;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 SLDKSTAVIPGYPVAGQ 20
DB 4 SLDKSDAXDLPRGPASGV 22

RESULT 4
T01765
endopeptidase La-like proteinase (EC 3.4.21.-) precursor, mitochondrial - Arabidopsis th
N:Alternate names: ATP-dependent proteinase LON; hypothetical protein A_IC002P16.23
N:Contents: adenosinetriphosphatase (EC 3.6.1.3)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 02-Jun-2003
C:Accession: T01765
R:Miller, N.; Beck, C.; Kramer, J.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of A. thaliana IG002P16.
A:Reference number: Z14421
A:Accession: T01765
A>Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1096 <MIL>
A:Cross-references: EMBL:AF007270; NID:g2191157; PIDN:AA61060.1; PID:g2191174; GSPDB:GN
C:Genetics:
A:Gene: ATP8A_IG002P16.23
A:Map position: 5
A:Introns: 99/3; 133/3; 155/3; 217/3; 239/3; 253/3; 296/1; 311/2; 329/3; 351/3; 398/3; 4
C:Function:
A:Description: serine proteinase
C:Superfamily: ATP-dependent Lon protease
C:Keywords: ATP; DNA binding; hydrolase; mitochondrial matrix; mitochondrion; molecular
F:567-574/Region: nucleotide-binding motif A (P-loop)
F:630-635/Region: nucleotide-binding motif B
F:573/Binding site: ATP (lys) #status predicted
F:982/Active site: Ser #status predicted

Query Match 46.9%; Score 46; DB 1; Length 1096;
Best Local Similarity 60.0%; Pred. No. 33;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VSDLKSTAVIPGYP 15
DB 1 VSDLKSTAVIPGYP 15

```

```

Db 78 VSDSKSSSAIVPTNP 92

RESULT 5
B65005
hypothetical protein b2324 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: B65005
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Ross, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: B65005
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-668 <BLAT>
A:Cross-references: GB:AB000321; GB:U00096; NID:gl788659; PIDN:AAK75384.1; PID:gl788664;
A:Experimental source: strain K-12, substrain MGL655

Query Match 45.9%; Score 45; DB 2; Length 668;
Best Local Similarity 60.0%; Pred. No. 29;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 6 SSTAIVPGYPVAGQV 20
DB 489 SQTSTLPVYSVAGQV 503

RESULT 6
AG1978
hypothetical protein alr1378 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG1978
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1978
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-779 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAK73335.1; PID:gl7130725; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1378

Query Match 45.9%; Score 45; DB 2; Length 779;
Best Local Similarity 44.4%; Pred. No. 33;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 DLKSTAVIPGYPVAGQV 20
DB 664 DSKSPTAIVPGIVIAQM 681

RESULT 7
AC2009
serine/threonine kinase with two-component sensor domain all1625 [imported] - Nostoc sp.
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AC2009
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2009

```



```

internalin proteins, probable peptidoglycan bound protein (LPXTG motif) homolog lmo2396
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AD1374
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; PMID:11679669
A:Accession: AD1374
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-940 <GLA>
A:Cross-references: GB:NC 003210; PIDN:CAD00474.1; PID:gl6411884; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2396

Query Match 44.4%; Score 43.5; DB 2; Length 940;
Best Local Similarity 50.0%; Pred. No. 72;
Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 1 VSDLKSTAVIPGYPVAGQV 20
DB 65 VADL-SNMSGAPGYPTGLI 83

RESULT 13
E97180
adenine phosphoribosyltransferase, Apt [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: E97180
R:Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A:Reference number: A96900; PMID:21359325; PMID:21359325
A:Accession: E97180
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-172 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80232.1; PID:gl5025279; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2275
C:Superfamily: adenine phosphoribosyltransferase.

Query Match 43.9%; Score 43; DB 2; Length 172;
Best Local Similarity 56.2%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 DLKSTAVIPGYPVAG 18
DB 2 DLKDSRVIDGPPKES 17

RESULT 14
H86939
probable ribose-phosphate pyrophosphokinase [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: H86939
R:Coile, S.T.; Eigimeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; PMID:21128732; PMID:11234002

```

```

A:Accession: H86939
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-327 <STO>
A:Cross-references: GB:AL450380; NID:gl3092591; PIDN:CAC29756.1; GSPDB:GN00147
C:Genetics:
A:Gene: prsA
C:Superfamily: ribose-phosphate pyrophosphokinase catalytic chain

Query Match 43.9%; Score 43; DB 2; Length 327;
Best Local Similarity 66.7%; Pred. No. 28;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 KSSTAVIPGYPVAGQ 19
DB 92 KEITAVIPFPYVARQ 106

RESULT 15
AD1456
hypothetical protein lin0187 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AD1456
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; PMID:21537279; PMID:11679669
A:Accession: AD1456
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-476 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC35420.1; PID:gl6412606; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin0187

Query Match 43.9%; Score 43; DB 2; Length 476;
Best Local Similarity 43.8%; Pred. No. 42;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 LKSTAVIPGYPVAGQ 19
DB 452 IKKNEIISGYPVGGK 467

Search completed: March 10, 2004, 12:12:30
Job time : 11.7692 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:51:36 ; Search time 6.15385 Seconds
(without alignments)
169.228 Million cell updates/sec

Title: US-10-044-703-69

Perfect score: 98

Sequence: 1 VSDLKSTAVIPGPVAGQV 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	100.0	159	MP63_MYCTU	P97175 mycobacteri
2	46.5	47.4	545	SGPL_DROME	Q9v7y2 drosophila
3	46	46.9	179	APT_BRAJA	Q89ab5 brachyrihob
4	46	46.9	940	LON2_ARATH	P93655 arabidopsis
5	45	45.9	668	YFCK_ECOLI	P77182 escherichia
6	45	45.9	668	YFCK_SHIFL	Q83qg8 shigella fl
7	44	44.9	428	DTAL_RALSO	Q8y2k5 ralstonia s
8	44	44.9	456	MYER_RAT	Q63k34 rattus norv
9	44	44.9	466	MYSR_MOUSE	O70342 mus musculu
10	44	44.9	1822	ZAP3_HUMAN	P49750 homo sapien
11	43	43.9	170	APT_BACAA	O81111 bacillus an
12	43	43.9	170	APT_BACCR	O817x3 bacillus ce
13	43	43.9	172	APT_CLOAB	Q97gu0 clostridium
14	43	43.9	327	KPR5_MYCLE	Q9cd45 mycobacteri
15	43	43.9	571	ATKA_MYCTU	P96371 mycobacteri
16	43	43.9	592	YTRP_PSEFU	P40604 pseudomonas
17	43	43.9	668	YFCK_EC057	Q8xqg7 escherichia
18	43	43.9	668	YFCK_EC016	Q8ifh0 escherichia
19	42	42.9	170	APT_BACSU	O34443 bacillus su
20	42	42.9	172	APT_LACPL	Q88vh0 lactobacill
21	42	42.9	263	TRPA_RHOSH	Q9x4e8 rhodobacter
22	42	42.9	267	M4AC_HUMAN	Q9xj10 homo sapien
23	42	42.9	401	UXUA_BRUNE	Q8ycd4 bruceella me
24	42	42.9	401	UXUA_BRUSU	Q8ifm2 bruceella su
25	42	42.9	471	GATA_THETH	Q91cx3 thermus the
26	42	42.9	520	LAC1_TRAHI	O02497 trametetes hi
27	42	42.9	520	LAC1_TRAVI	Q99044 trametetes vi
28	42	42.9	751	TAU_RAT	P19332 rattus norv
29	42	42.9	862	SNP2_YEAST	P32567 saccharomyc
30	42	42.9	1106	DP0D_BOVIN	P28339 bos taurus
31	42	42.9	1436	BRP5_MOUSE	Q9rix5 mus musculu
32	42	42.9	1518	BCA4_ACEXY	Q9rbj2 acetobacter
33	42	42.9	1518	BCA5_ACEXY	Q9wx75 acetobacter

ALIGNMENTS

RESULT 1

MP63_MYCTU
ID MP63_MYCTU STANDARD; PRT; 159 AA.
AC P97175; O08224;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Immunogenic protein MPT63/MPB63 precursor (Antigen MPT63/MPB63) (16
kDa immunoprotective extracellular protein).
DE MPT63 OR MPB63 OR RV1926C OR MT1377 OR MTCY09F9.38 OR MB1961C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=97130011; PubMed=8975987;
RA Manca C.M.A., Lyashchenko K., Wiker H.G., Usai D., Colangeli R.,
RA Gennaro M.L.;
RT "Molecular cloning, purification, and serological characterization of
MPT63, a novel antigen secreted by Mycobacterium tuberculosis";
RL Infect. Immun. 65:16-23(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=Erdmann;
RX MEDLINE=97313166; PubMed=9169770;
RA Harth G., Lee B.Y., Horwitz M.A.;
RT "High-level heterologous expression and secretion in rapidly growing
nonpathogenic mycobacteria of four major Mycobacterium tuberculosis
extracellular proteins considered to be leading vaccine candidates
and drug targets";
RL Infect. Immun. 65:2321-2328(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala E.,
Badock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornesby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,

34 41 41.8 142 1 Y441_MYCTU
35 41 41.8 187 1 APT_YERPE
36 41 41.8 292 1 ALF_STRP3
37 41 41.8 292 1 ALF_STRPY
38 41 41.8 298 1 SDB1_HUMAN
39 41 41.8 318 1 KPR2_YEAST
40 41 41.8 347 1 TRUD_VIBPA
41 41 41.8 371 1 YN06_SYNEL
42 41 41.8 395 1 RRPPI_P12H
43 41 41.8 395 1 RRPPI_P12HT
44 41 41.8 647 1 DPG1_CHICK
45 41 41.8 779 1 LEU2_YEAST
Q50813 mycobacteri
Q82c94 yersinia pe
Q8k5w5 streptococc
P82486 streptococc
O00560 homo sapien
P3620 saccharomyc
Q871q4 vibrio para
Q8dgl0 synecococc
P23055 human parai
P23056 human parai
Q92076 gallus gall
P07264 saccharomyc


```
CC phosphoribosyltransferase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF005944; BAC47755.1; -.
DR HAMAP; MF_00004; -; 1.
DR InterPro; IPR002375; PrPY_rptase.
DR InterPro; IPR008036; PRTtransferase.
DR Pfam; PF00105; Pribo syltran; 1.
DR PROSITE; PS00103; PUR_PVR_PR_TRANSFER; 1.
DR Transferase; Glycosyltransferase; Purine salvage; Complete proteome.
DR KN SEQUENCE 179 AA; 19405 MW; A7B600E0D869A61D CRG64;
DR SQ -----
Query Match 46.9%; Score 46; DB 1; Length 179;
Best Local Similarity 56.2%; Pred. No. 2.7;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 DLKSTAVIPGYVPVAG 18
DB 6 DLKASVRTIPDPKPG 21
|||||
|||

RESULT 4
LON2 ARATH
ID LON2 ARATH STANDARD; PRT; 940 AA.
AC P3665; O04954; QBRWX1;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-MAR-2004 (Rel. 43, Last annotation update)
DE Lon protease homolog 2, mitochondrial precursor (EC 3.4.21.-).
GN AT5G3686 OR F2P16.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosoids II; Brassicales; Brassicaceae; Arabidopsi s.
OX NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RC Sarria R., Lyznai k A., Mackenzie S.;
RT "Mitochondrial LON protease homolog from Arabidopsis thaliana
RL (Columbia)";
RN Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RC MEDLINE=21016721; PubMed=11130714;
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asanizu E.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohazaki N., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj D., O'Shaughnessy A., Preston J.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA Balter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozarsky P., Riley A., Strommatt C.,
RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,
RA Farnell L., Shah R., Rodriguez M., Hoan See L., Vil D., Baker J.,
RA Kirchhoff K., Toth K., King L., Bakret A., Miller B., Marra M.A.,
RA Hartnussen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Voickaert G., Wambutt R., Duescherhoef A., Stieckema W., Pohl T.,
RA Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,
RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RA Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,
RA van Staaveren M., Dirkse W., Moolliman P., Klein Lankhorst R.,
```

RA Weitzenecker T., Bothe G., Rose M., Hauf J., Berner S., Hempel S.,
RA Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,
RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rüd S., Schoof H.,
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.F.,
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
thaliana".
RL Nature 408:823-826(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Hevan V.W., Iida K., Karnes M.,
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA Kamiya A., Meyers C., Nakajima M., Naruoka M., Seki M., Sakurai T.,
RA Satou M., Tameze R., Vayenberg M., Wallender E.K., Wong C., Yamamura Y.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.,
RT "Empirical analysis of transcriptional activity in the Arabidopsis
genome".
RL Science 302:842-846(2003).
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- SIMILARITY: Belongs to peptidase family S16.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
in positions 718, 724 and 726.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U88087; AAB48000.1; ALT FRAME.
DR EMBL: AF007270; AAB61060.1; ALT_SEQ.
DR EMBL: AF091049; AAM13870.1; -.
DR EMBL: A1117355; AAM51430.1; -.
DR InterPro: IPR003593; AAA ATPase.
DR InterPro: IPR003959; AAA ATPase centr.
DR InterPro: IPR008269; Pept S16 C.
DR InterPro: IPR003111; Pept S16 N.
DR InterPro: IPR008268; Peptid S16 AS.
DR InterPro: IPR001984; Peptidase_S16.
DR Pfam: PF00004; AAA; 1.
DR Pfam: PF02190; LON; 1.
DR Pfam: PF05362; Lon C; 1.
DR PRINTS: PR00830; ENDOLAPTASE.
DR SMART: SM00382; AAA; 1.
DR SMART: SM00464; LON; 1.
DR PROSITE: PS01046; LON SER; 1.
KW Hydrolase; Serine protease; ATP-binding; Multigene family;
KW Mitochondrion; Transit peptide.
FT TRANSIT 1 61 MITOCHONDRION (POTENTIAL).
FT CHAIN 62 940 LON PROTEASE HOMOLOG 2.
FT NP BIND 464 471 ATP (POTENTIAL).
FT ACT SITE 841 841 BY SIMILARITY.
FT CONFLICT 288 288 V -> F (IN REF. 1).
FT CONFLICT 500 500 K -> Q (IN REF. 1).
FT CONFLICT 679 679 A -> S (IN REF. 1).
FT CONFLICT 935 936 YD -> FA (IN REF. 1).
SQ SEQUENCE 940 AA; 103929 MW; 6060A8BC9AE5E28 CRC64;

Query Match 46.9%; Score 46; DB 1; Length 940;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VSDLSKSTAVIPGYP 15
Db 78 VSDSKSSSAIVFTNP 92
RESULT 5
YFCK_ECOLI STANDARD; PRT; 568 AA.
ID YFCK_ECOLI
AC P77182;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical UPF0209 protein yfck.
GN YFCK OR B2324.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R.; Plunkett G. III; Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Sato Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features".
RL DNA Res. 4:91-113(1997).
RN -!- SIMILARITY: Belongs to the UPF0209 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000321; AAC75384.1; ALT_INIT.
DR EMBL: D90863; BAA16181.1; ALT_INIT.
DR EcoGene; EGI4114; Yfck.
DR HAMAP; MF 01102; -; 1.
DR InterPro: IPR008471; DUF752.
DR Pfam: PF05430; DUF752; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 568 AA; 74434 MW; DD24632C5B6971B CRC64;
Query Match 45.9%; Score 45; DB 1; Length 668;
Best Local Similarity 60.0%; Pred. No. 17;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 6 SSTAVIPGYPVAGQV 20
Db 469 SQTSTLPVSVAGQV 483
RESULT 6
YFCK_SHIFL

RA Gerald C., Walker M.W., Criscione L., Gustafson E.L.,
RA Batzi-Hartmann C., Smith K.E., Vayse P., Durkin M.M., Laz T.M.,
RA Linemeyer D.L., Schaffhauser A.O., Whitebread S., Hofbauer K.G.,
RA Taber R.I., Branchek T.A., Weinshank R.L.;
RT "A receptor subtype involved in neurotrophin-induced food intake.";
RL Nature 382:168-171(1996).
RN [2]
RP SEQUENCE OF 12-456 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98332165; PubMed=9669502;
RA Parker E.M., Babji C.K., Balasubramanian A., Burtner R.E., Guzzi M.,
RA Hamud F., Mukhopadhyay G., Rudinski M.S., Tao Z., Tice M., Xia L.,
RA Mullins D.E., Salisbury B.G.;
RT "GR231118 (1229U91) and other analogues of the C-terminus of
RT neuropeptide Y are potent neuropeptide Y Y1 receptor antagonists and
RT neuropeptide Y Y4 receptor agonists.";
RL Eur. J. Pharmacol. 349:97-105(1998).
RN [3]
RP SEQUENCE OF 12-456 FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=96421636; PubMed=8824284;
RA Hu Y., Bloomquist B.T., Cornfield L.J., Decarr L.B.,
RA Flores-Riveros J.R., Friedman L., Jiang P., Lewis-Higgins L.,
RA Sadowski Y., Schaefer J., Velazquez N., McCabe M.L.;
RT "Identification of a novel hypothalamic neuropeptide Y receptor
RT associated with feeding behavior.";
RL J. Biol. Chem. 271:26315-26319(1996).
CC -!- FUNCTION: Receptor for neuropeptide Y and peptide YY. The activity
CC of this receptor is mediated by G proteins that inhibit adenylylate
CC cyclase activity. Seems to be associated with food intake. Could
CC be involved in feeding disorders.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Brain; hypothalamus.
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC Highest to tachykinins receptors.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U56078; AAC52677.1; -;
CC EMBL: AF044264; AAC18670.1; -;
CC EMBL: U66274; AAC52845.1; -;
CC InterPro: IPR000276; GPCR_Rhodpsn.
CC Pfam: PF00001; 7tm_1; 1.
CC PRINTS: PR00237; GPCR_Rhodpsn.
CC PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; FALSE_NEG.
CC PROSITE: PS00262; G_PROTEIN_RECEP_F1_2; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Phosphorylation; Lipoprotein; Palmitate.
FT DOMAIN 1 51
FT TRANSMEM 52 73
FT DOMAIN 74 85
FT TRANSMEM 86 106
FT DOMAIN 107 126
FT TRANSMEM 127 148
FT DOMAIN 149 168
FT TRANSMEM 169 189
FT DOMAIN 190 221
FT TRANSMEM 222 243
FT DOMAIN 244 378
FT TRANSMEM 379 401
FT DOMAIN 402 414
FT TRANSMEM 415 438
FT DOMAIN 439 456
FT DISULFID 125 209
FT LIPID 452 452
FT CARBOHYD 21 21
FT CARBOHYD 28 28

SQ SEQUENCE 456 AA; 51715 MW; 5157ABE341BA707E CRC64;
Query Match 44.9%; Score 44; DB 1; Length 456;
Best Local Similarity 47.4%; Pred. No. 16;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
QY 2 SDKSSTAVIPGVPVAGQV 20
DB 331 SQLSPSSKVIQVPICFEV 349
RESULT 9
NY5R MOUSE STANDARD; PRT; 466 AA.
AC 070342; O35380; O9JMK1;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neuropeptide Y receptor type 5 (NPY5-R) (NPY-5 receptor) (Y5
DE receptor).
GN NPY5R OR NPY5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99017379; PubMed=9802393;
RA Borowsky B., Walker M.W., Bard J., Weinshank R.L., Laz T.M.,
RA Vayse P., Branchek T.A., Gerald C.;
RT "Molecular biology and pharmacology of multiple NPY Y5 receptor
RT species homologs.";
RL Regul. Pept. 75:45-53(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/Sv; TISSUE=Brain;
RA Chen H., Adams S., McWhinnie E., Bayne M., Gadske R., Zastawny R.;
RT "Mouse neuropeptide Y Y5 receptor characterized by repeat sequence in
RT extracellular domain.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBSJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97459646; PubMed=9315606;
RA Nakamura M., Yokoyama M., Watanabe H., Matsumoto T.;
RT "Molecular cloning, organization and localization of the gene for the
RT mouse neuropeptide Y-Y5 receptor.";
RL Biochim. Biophys. Acta 1328:83-89(1997).
CC -!- FUNCTION: Receptor for neuropeptide Y and peptide YY. The activity
CC of this receptor is mediated by G proteins that inhibit adenylylate
CC cyclase activity. Seems to be associated with food intake. Could
CC be involved in feeding disorders (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC Highest to tachykinins receptors.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF049329; AAC17839.1; -;
CC EMBL: AF022948; AAB81829.1; -;
CC EMBL: AB001346; BAA89538.1; -;
CC MGD: MGI:108082; Npy5r.
CC GO: GO:0016020; C-membrane; IDA.
CC DR GO: GO:0001602; F-pancreatic polypeptide receptor activity; IDA.
CC DR GO: GO:0001601; Peptide YY receptor activity; IDA.
CC DR GO: GO:0007273; P-regulation of synapse; IMP.
CC InterPro: IPR000276; GPCR_Rhodpsn.
CC Pfam: PF00001; 7tm_1; 1.

```

DR DR PROSITE; PS00237; GPCRHHODSPS.
DR DR PROSITE; PS00237; G PROTEIN RECEPTOR F1.1; FALSE_NEG.
DR DR PROSITE; PS00262; G PROTEIN RECEPTOR F1.2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW phosphorylation; Lipoprotein; Palmitate.
FT DOMAIN 1 61 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 62 83 1 (POTENTIAL).
FT DOMAIN 84 95 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 96 116 2 (POTENTIAL).
FT DOMAIN 117 136 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 137 158 3 (POTENTIAL).
FT DOMAIN 159 178 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 179 199 4 (POTENTIAL).
FT DOMAIN 200 231 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 232 253 5 (POTENTIAL).
FT DOMAIN 254 388 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 389 411 6 (POTENTIAL).
FT DOMAIN 412 424 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 425 448 7 (POTENTIAL).
FT DOMAIN 449 466 CYTOPLASMIC (POTENTIAL).
FT DISULFID 467 486 BY SIMILARITY.
FT LIPID 462 462 S-palmitoyl cysteine (POTENTIAL).
FT CARBOHYD 10 10 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 17 17 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 39 39 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 195 195 L -> F (IN REF. 1).
FT CONFLICT 284 284 K -> Q (IN REF. 3).
SQ SEQUENCE 466 AA; 52784 MW; B157F236EF2D4385 CRC64;

Query Match 44.9%; Score 44; DB 1; Length 466;
Best Local Similarity 47.4%; Pred. No. 16;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 SDLKSTAVIPGVVAGQV 20
DB 341 SQLSPSSKVIQGVICREV 359

RESULT 10
ZAP3_HUMAN
ID ZAP3_HUMAN STANDARD; PRT; 1822 AA.
AC P49750; P49752; Q9PIV7;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear protein ZAP3 (ZAP113).
GN ZAP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Madan A., Qin S., Abbasi N., Baradarani L., Birditt B.,
RA Bloom S., Dors M., Dickhoff R., Fleetwood P., Harrison G., James R.,
RA Kaur A., Madan A., Owen M.P., Ratcliffe A., Shaifer T., Hood L.;
RT "Sequencing of human chromosome 14q24.3 region."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 539-847 AND 1397-1822 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95319502; PubMed=7596406;
RA Sherrington R., Rogaeve E.I., Liang Y., Rogaeve E.A., Levesque G.,
RA Ikeda M., Chi H., Lin C., Li G., Holman K., Tsuda T., Mar L.,
RA Foncin J.-F., Bruni A.C., Montesi M.P., Sorbi S., Rainero I.,
RA Pinessi L., Nee L., Chumakov I., Pollen D., Brookes A.,
RA Sanseau P., Polinsky R.J., Wasco W., da Silva H.A.R., Haines J.L.,
RA Pericak-Vance M.A., Tanzi R.E., Roses A.D., Fraser P.E.,
RA Rommens J.M., St George-Hyslop P.H.;
RT "Cloning of a gene bearing missense mutations in early-onset familial
RT Alzheimer's disease."
RL Nature 375:754-760(1995).

```

```

CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1661.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AC007956; AAF61275.1;
CC EMBL; L40403; AAC42008.1; ALT_FRAME.
CC EMBL; L40400; AAC42006.1;
CC Nuclear protein.
KW DOMAIN 15 205 PRO-RICH.
FT DOMAIN 382 430 GLN-RICH.
FT DOMAIN 807 1209 ARG-RICH.
FT DOMAIN 1488 1577 ARG-RICH.
FT CONFLICT 621 621 P -> S (IN REF. 2).
FT CONFLICT 1404 1404 T -> I (IN REF. 2).
FT CONFLICT 1821 1821 K -> E (IN REF. 2).
SQ SEQUENCE 1822 AA; 204947 MW; 8E6CB83FE540C7D2 CRC64;

Query Match 44.9%; Score 44; DB 1; Length 1822;
Best Local Similarity 60.0%; Pred. No. 72;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 SDLKSTAVIPGVV 16
DB 1398 SEFKSETAIPSAVP 1412

RESULT 11
APT_BACAA
ID APT_BACAA STANDARD; PRT; 170 AA.
AC Q81L11;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT).
GN APT OR BA4638.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Rikstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Benton J.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria."
RL Nature 423:81-86(2003).
CC -!- FUNCTION: Catalyzes a salvage reaction resulting in the formation
CC of AMP, that is energetically less costly than de novo synthesis.
CC -!- CATALYTIC ACTIVITY: AMP + diphosphate = adenine + 5-phospho-alpha-
CC D-ribose 1-diphosphate.
CC -!- PATHWAY: Purine salvage.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the purine/pyrimidine
CC phosphoribosyltransferase family.
CC -----

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE017039; AAP28341.1; --
 DR TIGR; BA4638; --
 DR HAMAP; MF 00004; --; 1.
 DR InterPro; IPR005764; Ade_phospho_trans.
 DR InterPro; IPR002375; Pr/py_rp_transf.
 DR InterPro; IPR000836; Prtransferase.
 DR Pfam; PF00156; Pribosyltran; 1.
 DR TIGRFAMs; TIGR01090; apt; 1.
 DR PROSITE; PS00103; PUR_PYR_PR_TRANSFER; FALSE NEG.
 DR TRANSFERASE; Glycosyltransferase; Purine salvage; Complete proteome.
 DR KW TRANSFERASE; Glycosyltransferase; Purine salvage; Complete proteome.
 DR SQ SEQUENCE 170 AA; 18644 MW; 148BA98F47CF8FFC CRC64;
 Query Match 43.9%; Score 43; DB 1; Length 170;
 Best Local Similarity 44.4%; Pred. No. 8.1;
 Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
 QY 3 DLKSTAVIPGYPVAGQV 20
 DB 2 DFKQHIATVPDYPKEGIV 19
 RESULT 12
 APT_BACOR
 ID APT_BACOR STANDARD; PRT; 170 AA.
 AC Q817X3;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT).
 GN APT OR BC4402.
 OS Bacillus cereus (strain ATCC 14579 / DSM 31).
 CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 CC NCBI_TaxID=226900;
 RN [1]
 RP MEDLINE=22608415; PubMed=12721630;
 RA Ivarova N., Sorokin A., Anderson I., Galleron N., Candelson B.,
 RA Kapral V., Shattacharya A., Renik G., Mikhailova N., Lapidus A.,
 RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
 RA Grechkin Y., Pusch G., Haselkorn R., Feinstein M., Ehrlich S.D.,
 RA Overbeek R., Kyrpides N.;
 RT "Genome sequence of Bacillus cereus and comparative analysis with
 Bacillus anthracis."
 RL Nature 423:87-91(2003).
 CC -!- FUNCTION: Catalyzes a salvage reaction resulting in the formation
 CC of AMP, that is energetically less costly than de novo synthesis.
 CC -!- CATALYTIC ACTIVITY: AMP + diphosphate = adenine + 5-phospho-alpha-
 CC D-ribose 1-diphosphate.
 CC -!- PATHWAY: Purine salvage.
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the purine/pyrimidine
 CC phosphoribosyltransferase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; AE017012; AAP11315.1; --
 DR HAMAP; MF 00004; --; 1.
 DR InterPro; IPR005764; Ade_phospho_trans.

DR InterPro; IPR002375; Pr/py_rp_transf.
 DR InterPro; IPR000836; Prtransferase.
 DR Pfam; PF00156; Pribosyltran; 1.
 DR TIGRFAMs; TIGR01090; apt; 1.
 DR PROSITE; PS00103; PUR_PYR_PR_TRANSFER; FALSE NEG.
 DR TRANSFERASE; Glycosyltransferase; Purine salvage; Complete proteome.
 DR KW TRANSFERASE; Glycosyltransferase; Purine salvage; Complete proteome.
 DR SQ SEQUENCE 170 AA; 18630 MW; A491A98F47D58FF7 CRC64;
 Query Match 43.9%; Score 43; DB 1; Length 170;
 Best Local Similarity 44.4%; Pred. No. 8.1;
 Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
 QY 3 DLKSTAVIPGYPVAGQV 20
 DB 2 DFKQHIATVPDYPKEGIV 19
 RESULT 13
 APT_CLOAB
 ID APT_CLOAB STANDARD; PRT; 172 AA.
 AC Q97GU0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT).
 GN APT OR CAC2275.
 OS Clostridium acetobutylicum.
 CC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 CC Clostridium.
 CC NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RA MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Boret G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 Clostridium acetobutylicum."
 RL J. Bacteriol. 183:4823-4838(2001).
 CC -!- FUNCTION: Catalyzes a salvage reaction resulting in the formation
 CC of AMP, that is energetically less costly than de novo synthesis.
 CC -!- CATALYTIC ACTIVITY: AMP + diphosphate = adenine + 5-phospho-
 CC alpha-D-ribose 1-diphosphate.
 CC -!- PATHWAY: Purine salvage.
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to the purine/pyrimidine
 CC phosphoribosyltransferase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; AE007728; AAK80232.1; --
 DR FRL; E97180; E97180.
 DR HAMAP; MF 00004; --; 1.
 DR InterPro; IPR005764; Ade_phospho_trans.
 DR InterPro; IPR002375; Pr/py_rp_transf.
 DR InterPro; IPR000836; Prtransferase.
 DR Pfam; PF00156; Pribosyltran; 1.
 DR TIGRFAMs; TIGR01090; apt; 1.
 DR PROSITE; PS00103; PUR_PYR_PR_TRANSFER; FALSE NEG.
 DR TRANSFERASE; Glycosyltransferase; Purine salvage; Complete proteome.
 DR KW TRANSFERASE; Glycosyltransferase; Purine salvage; Complete proteome.
 DR SQ SEQUENCE 172 AA; 18855 MW; F6B7B0EE32FE631 CRC64;
 Query Match 43.9%; Score 43; DB 1; Length 172;
 Best Local Similarity 56.2%; Pred. No. 8.2;


```
CC -!- SIMILARITY: Belongs to the kdpA family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z92539; CAB06850.1; -
DR EMBL; AE006988; AAK45310.1; -
DR EMBL; BX248337; CAD93918.1; -
DR PIR; H70623; H70623.
DR TIGR; MT1058; -
DR TubercuList; RV1029; -
DR HAMAP; MF_00275; -; 1.
DR InterPro; IPR004623; K_ArPaseA.
DR Pfam; PF03814; KdpA; 1.
DR TIGRFAMs; TIGR00680; kdpA; 1.
KW Hydrolase; Transport; Potassium transport; Transmembrane;
KW Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 66 86 POTENTIAL.
FT TRANSMEM 137 157 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
FT TRANSMEM 255 275 POTENTIAL.
FT TRANSMEM 286 306 POTENTIAL.
FT TRANSMEM 369 389 POTENTIAL.
FT TRANSMEM 390 410 POTENTIAL.
FT TRANSMEM 430 450 POTENTIAL.
FT TRANSMEM 497 517 POTENTIAL.
FT TRANSMEM 539 559 POTENTIAL.
SQ SEQUENCE 571 AA; 60163 MW; 4A3238D1C62A3687 CRC64;

Query Match 43.9%; Score 43; DB 1; Length 571;
Best Local Similarity 52.9%; Pred No; 30;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 LKSSTAVIPGYPVAGQV 20
Db 212 LEGAPQLIPGPFVAGQV 228
| : : ||| |||
| : : ||| |||
```

Search completed: March 10, 2004, 12:06:33
Job time : 7.15385 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:57:36 ; Search time 31.6667 Seconds
(without alignments)
199.275 Million cell updates/sec

Title: US-10-044-703-69

Perfect score: 98

Sequence: 1 VSLKSTAVIPGYPVAGQV 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	51.0	314	16 Q9KAQ3	Q9kaq3 bacillus ha
2	50	51.0	408	9 Q7Y3F6	Q7y3f6 streptococ
3	49	50.0	396	17 Q8PVN1	Q8pvn1 methanosarc
4	47	48.0	834	16 Q82UM9	Q82um9 nitrosomona
5	46.5	47.4	545	5 Q9V7Y2	Q9v7y2 drosophila
6	46	46.9	142	5 Q8SV15	Q8sv15 enccephalito
7	46	46.9	179	16 Q89SB5	Q89sb5 bradyrhizob
8	46	46.9	312	16 Q9A7V9	Q9a7v9 caulobacter
9	46	46.9	397	17 Q8TIQ5	Q8tiq5 methanosarc
10	46	46.9	680	16 Q81B19	Q81b19 bacillus ce
11	45	45.9	147	5 Q962D5	Q962d5 drosophila
12	45	45.9	147	5 Q81N42	Q81n42 drosophila
13	45	45.9	293	2 Q9FA98	Q9fa98 streptococ
14	45	45.9	501	16 Q7UW43	Q7uw43 rhodospirell
15	45	45.9	573	10 Q94D10	Q94d10 oryza sativ
16	45	45.9	688	16 Q83QQ8	Q83qq8 shigella fl

17	45	45.9	779	16 Q8YX39	Q8yx39 anabaena sp
18	45	45.9	1787	16 Q8YWI8	Q8ywi8 anabaena sp
19	44	44.9	263	10 Q8LG33	Q8lg33 arabidopsis
20	44	44.9	263	10 Q9C773	Q9c773 arabidopsis
21	44	44.9	265	2 Q44134	Q44134 actinobacil
22	44	44.9	309	5 Q8IMV1	Q8imv1 drosophila
23	44	44.9	329	5 Q8SY52	Q8sy52 drosophila
24	44	44.9	339	10 Q941I0	Q941i0 fragaria an
25	44	44.9	349	10 Q9FWR7	Q9fw7 arabidopsis
26	44	44.9	357	17 Q9V0G4	Q9v0g4 pyrococcus
27	44	44.9	424	16 Q7UPN6	Q7upn6 rhodopirell
28	44	44.9	446	10 Q9FTH0	Q9fti0 oryza sativ
29	44	44.9	694	16 Q98FE1	Q98pe1 mycoplasma
30	44	44.9	843	16 Q7W3Z1	Q7w3z1 bordetella
31	44	44.9	843	16 Q7W9R7	Q7w9r7 bordetella
32	44	44.9	847	16 Q7VMA9	Q7vma9 bordetella
33	44	44.9	1200	10 Q7XK75	Q7xk75 oryza sativ
34	44	44.9	1604	10 Q9X140	Q9x140 arabidopsis
35	44	44.9	1766	4 Q8NF45	Q8nf45 homo sapien
36	43.5	44.4	940	16 Q8Y4N9	Q8y4n9 listeria mo
37	43	43.9	198	10 Q9LJ29	Q9lj29 oryza sativ
38	43	43.9	293	10 Q93ZB2	Q93zb2 arabidopsis
39	43	43.9	322	10 Q84V25	Q84v25 fragaria an
40	43	43.9	457	10 Q7X618	Q7x618 oryza sativ
41	43	43.9	476	16 Q92FB9	Q92fb9 listeria in
42	43	43.9	509	10 Q81188	Q81188 arabidopsis
43	43	43.9	509	10 Q9S7C8	Q9s7c8 arabidopsis
44	43	43.9	585	10 Q9LJD8	Q9lyd8 arabidopsis
45	43	43.9	662	16 Q8F3Y2	Q8fy2 leptospira

ALIGNMENTS

RESULT 1

Q9KAQ3 PRELIMINARY; PRT; 314 AA.
ID Q9KAQ3;
AC Q9KAQ3;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Response regulatory protein (sensory transduction system).
GN BH2234.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]_TaxID=86665;
RP SEQUENCE FROM N.A.
RC STRAIN=C-325 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasore K., Takaki Y., Maeno G., Sasaki R., Masui N.,
Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
Horikoshi K.;
RA "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.
EMBL; AF001514; BAB05953.1; -
DR PIR; B83929; B83929.
DR HSSP; P08402; 1B00.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0000156; P:two-component response regulator activity; IEA.
DR GO; GO:0007600; P:sensory perception; IEA.
DR GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR InterPro; IPR00160; GDEF.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF00990; GDEF; 1.
DR Pfam; PF00072; response_reg; 1.
DR Prodom; P000039; Response_reg; 1.
DR SMART; SM00267; DUF1; 1.
DR SMART; SM00448; REC; 1.
DR TIGRFAMs; TIGR00254; GDEF; 1.

DR PROSITE; PS50887; GDEF; 1.
 DR PROSITE; PS50110; RESPONSE REGULATORY; 1.
 KW Phosphorylation; Sensory transduction; Complete proteome.
 SQ SEQUENCE 314 AA; 36056 MW; 6A0B54E99B9506CD CRC64;

Query Match 51.0%; Score 50; DB 16; Length 314;
 Best Local Similarity 57.9%; Pred. No. 8.1;
 Matches 11; Conservative 1; Mismatches 0; Gaps 0;

QY 1 VSDLKSTAVIPGYPVAGQ 19
 Db 80 VSSLKEKEAVIKGEVGGQ 98

RESULT 2
 QY3F6 PRELIMINARY; PRT; 408 AA.
 ID QY3F6
 AC QY3F6;
 DT 01-OCT-2003 (TRENBLrel. 25, Created)
 DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Neck appendage.
 GN ORF6.
 OS Streptococcus phage C1.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
 OC phi-29-like viruses.
 CX NCBI_TaxID=230871;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22638289; PubMed=12754230;
 RA Nelson D., Schuch R., Zhu S., Tschern D.M., Fischetti V.A.;
 RT "Genomic Sequence of C1, the First Streptococcal Phage."
 RL J. Bacteriol. 185:3325-3332(2003).
 DR EMBL; AY212251; AAP42305.1; -
 SQ SEQUENCE 408 AA; 47805 MW; 18FA3AC07C205CB2 CRC64;

Query Match 51.0%; Score 50; DB 9; Length 408;
 Best Local Similarity 66.7%; Pred. No. 11;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 SLDKSTAVIPGYPV 16
 Db 142 SLDKSSAVLNYPV 156

RESULT 3
 Q8PYN1 PRELIMINARY; PRT; 396 AA.
 ID Q8PYN1
 AC Q8PYN1;
 DT 01-OCT-2002 (TRENBLrel. 22, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE Molybdopter in biosynthesis MoeA protein.
 GN MM0830.
 OS Methanosarcina maezi (Methanosarcina frisia).
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 CX NCBI_TaxID=2209;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=Goel / Go1 / ATCC BAA-199 / DSM 3647 / OCM 89;
 RC MEDLINE=22120827; PubMed=12125824;
 RA Deppenweier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
 RA Martinez-Arias R., Henne A., Wietz A., Baumer S., Jacobi C.,
 RA Bruggemann H., Lienard T., Christmann A., Boemcke M., Stachel S.,
 RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
 RA Fritz H.-J., Gottschalk G.;
 RT "The genome of Methanosarcina maezi: evidence for lateral gene
 transfer between Bacteria and Archaea."
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
 DR EMBL; AE013308; AAM30526.1; -
 DR GO; GO:0006777; P-Mo-molybdopter in cofactor biosynthesis; IBA.
 DR InterPro; IPR001453; MOCF_Biosynth.

DR InterPro; IPR005111; MoeA_C.
 DR InterPro; IPR005110; MoeA_N.
 DR Pfam; PF00994; MOCF_biosynth; 1.
 DR Pfam; PF03454; MoeA_C; 1.
 DR Pfam; PF03453; MoeA_N; 1.
 DR ProDom; PD002460; MOCF_biosynth; 1.
 DR TIGRFAMs; TIGR00177; molyb_syn; 1.
 KW Complete proteome.
 SQ SEQUENCE 396 AA; 42324 MW; F1B732CD7A9AF9EC CRC64;

Query Match 50.0%; Score 49; DB 17; Length 396;
 Best Local Similarity 56.2%; Pred. No. 15;
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 KSSTAVIPGYPVAGQV 20
 Db 289 KTVVCLPGYPVAGLV 304

RESULT 4
 Q82UM9 PRELIMINARY; PRT; 834 AA.
 ID Q82UM9
 AC Q82UM9;
 DT 01-JUN-2003 (TRENBLrel. 24, Created)
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE DUF214.
 GN NE1453.
 OS Nitrosomonas europaea.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
 OC Nitrosomonadaceae; Nitrosomonas.
 CX NCBI_TaxID=915;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=ATCC 19718 / IFO 14298;
 RC MEDLINE=22586410; PubMed=12700255;
 RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
 RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
 RA Arciero D.M., Holmes N.G., Whittaker M.M., Arp D.J.;
 RT "Complete genome sequence of the ammonia-oxidizing bacterium and
 RT obligate chemolithoautotroph Nitrosomonas europaea."
 RL J. Bacteriol. 185:2759-2773(2003).
 DR EMBL; BX321861; CAD85364.1; -
 DR GO; GO:0016020; C-membrane; IEA.
 DR InterPro; IPR003838; DUF214.
 DR Pfam; PF02687; FtsX; 2.
 KW Complete proteome.
 SQ SEQUENCE 834 AA; 92309 MW; 50B914650E554935 CRC64;

Query Match 48.0%; Score 47; DB 16; Length 834;
 Best Local Similarity 52.9%; Pred. No. 73;
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 LKSTAVIPGYPVAGQV 20
 Db 101 LTVSVAVAPGYPVLRGEL 117

RESULT 5
 Q9V7Y2 PRELIMINARY; PRT; 545 AA.
 ID Q9V7Y2
 AC Q9V7Y2;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE CG8946 protein (SPHINGOSINE-phosphate lyase) (SD02978P).
 GN SFLY OR SPL OR CG8946.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 CX NCBI_TaxID=7227;
 RN [1]


```

DR GO: GO:0009116; P:nucleoside metabolism; IEA.
DR InterPro: IPR002375; Fz/PY_rpf_transf.
DR InterPro: IPR000836; P:transf.
DR Pfam: PF00156; Pribosyltran; 1.
DR PROSITE: PS00103; PUR_PYR_PR_TRANSFER; 1.
DR Glycosyltransferase; Transferase; Complete proteome.
KW SEQUENCE 179 AA; 19405 MW; A7B600E0D869A61D CRC64;

Query Match 46.9%; Score 46; DB 16; Length 179;
Best Local Similarity 56.2%; Pred. No. 20;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 3 DLKSTAVIPGYPVAG 18
Db 6 DLKASVRTIPDPKPG 21

RESULT 8
Q9A7V9 PRELIMINARY; PRT; 312 AA.
AC Q9A7V9;
DT 01-JUN-2001 (TrEMBLrel. 17; Created)
DT 01-JUN-2001 (TrEMBLrel. 17; Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24; Last annotation update)
DE Hypothetical protein CC1610.
GN CC1610.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uitterback T., Iran K., Wolf A., Vamatheva J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005836; AAK23589.1; -.
DR PIR: A87449; A87449.
DR TIGR: CC1610; -.
DR InterPro: IPR007492; LytTR.
DR Pfam: PF04397; LytTR; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 312 AA; 34049 MW; 44DC3E2C327CF910 CRC64;

Query Match 46.9%; Score 46; DB 16; Length 312;
Best Local Similarity 52.6%; Pred. No. 37;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 SDLKSTAVIPGYPVAGV 20
Db 4 SDLKSDAKDLPRGPASGV 22

RESULT 9
Q8TIQ5 PRELIMINARY; PRT; 397 AA.
AC Q8TIQ5;
DT 01-JUN-2002 (TrEMBLrel. 21; Created)
DT 01-JUN-2002 (TrEMBLrel. 21; Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24; Last annotation update)
DE Molybdenum cofactor biosynthesis protein MoeA2.
GN MA4087.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=1932238;
RA Galign J.E., Nuebaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atncor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.B., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Unayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR EMBL; AE011121; AAM07435.1; -.
DR GO: GO:0006777; P:Mo-molybdopterin cofactor biosynthesis; IEA.
DR InterPro: IPR001453; MoCF_biosynth.
DR InterPro: IPR005111; MoeA_C.
DR InterPro: IPR005110; MoeA_N.
DR Pfam: PF00994; MoCF_biosynth; 1.
DR Pfam: PF03454; MoeA_C; 1.
DR Pfam: PF03453; MoeA_N; 1.
DR ProDom: PD002460; MoCF_biosynth; 1.
DR TIGRFAMs: TIGR00177; molyb_syn; 1.
KW Complete proteome.
SQ SEQUENCE 397 AA; 42438 MW; 0488E0E2778A2D0D CRC64;

Query Match 46.9%; Score 46; DB 17; Length 397;
Best Local Similarity 56.2%; Pred. No. 48;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 5 KSSTAVIPGYPVAGV 20
Db 289 KVPVCLPGYPVAGLV 304

RESULT 10
Q81B19 PRELIMINARY; PRT; 680 AA.
AC Q81B19;
DT 01-JUN-2003 (TrEMBLrel. 24; Created)
DT 01-JUN-2003 (TrEMBLrel. 24; Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25; Last annotation update)
DE Xanthine dehydrogenase molybdopterin-binding subunit (EC
DE 1.1.1.204).
GN BC3168.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapral V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyripides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis."
RL Nature 423:87-91(2003).
DR EMBL; AE017008; AAP10110.1; -.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0006118; P:electron transport; IEA.
DR InterPro: IPR008274; Aldxan_dh_bind.
DR Pfam: PF02738; Ald_xan_dh_C2; 1.
DR Oxidoreductase; Complete proteome.
KW SEQUENCE 680 AA; 74501 MW; 3171DF3FCF5E4FD4 CRC64;

Query Match 46.9%; Score 46; DB 16; Length 680;

```

```

Best Local Similarity 43.8%; Pred. No. 86;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VSDLKSTAVIPGPV 16
Db 447 ISQLKKSASILLGYP 462

RESULT 11
Q962D5 PRELIMINARY; PRT; 147 AA.
AC Q962D5;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Stress-inducible humoral factor Turandot 2.
GN TOTZ OR CG31507.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RA Ekengren S., Hultmark D.;
RT "A family of Turandot-related genes in the humoral stress response of
RL Drosophila.";
RL Biochem. Biophys. Res. Commun. 0:0-0(2001).
DR EMBL; AV035994; AAK64527.1; -.
DR FlyBase; FBgn0044809; TotZ.
SQ SEQUENCE 147 AA; 16285 MW; 44E09D5DA09C04B5 CRC64;

Query Match 45.9%; Score 45; DB 5; Length 147;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Qy 1 VSDLKSTAVIPGPVAGQV 20
Db 81 VNDFKRKTWIDGVPAGGV 100

RESULT 12
Q8IN42 PRELIMINARY; PRT; 147 AA.
AC Q8IN42;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE CG31507-PA.
GN TOTZ OR CG31507.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananatiides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers V.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Arif J.P., Agbayani A., An H.J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J.P., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

```

```

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glödek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Stupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Ananatiides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreria S., Frise E., Galie R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Parasas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003732; AANI1842.1; -.
DR FlyBase; FBgn0044809; TotZ.
SQ SEQUENCE 147 AA; 16179 MW; 450A92CEA5FF0E45 CRC64;

Query Match 45.9%; Score 45; DB 5; Length 147;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Qy 1 VSDLKSTAVIPGPVAGQV 20
Db 81 VNDFKRKTWIDGVPAGGV 100

```

RESULT 13

Q9FA38 PRELIMINARY; PRT; 293 AA.
AC Q9FA98;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Class-II aldolase.
GN FBA.
OS Streptococcus bovis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1315;
RN [1]_taxID=1315;
RP SEQUENCE FROM N.A.
RA Asanuma N., Hino T.;
RT "Streptococcus bovis fba gene for putative class-II aldolase, complete cds."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB050113; BAB16889.1; -.
DR HSSP; P11604; 1DCS.
DR GO; GO:0004332; F:fructose-bisphosphate aldolase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006036; P:glycolysis; IEA.
DR InterPro; IPR000771; K_bp_aldolase.
DR Pfam; PF01116; F_bp_aldolase; 1.
DR ProDom; PD023767; K_bp_aldolase; 1.
DR TrEMBL; TIGR00167; cbaA; 1.
DR PROSITE; PS00602; ALDOLASE_CLASS_II_1; 1.
DR PROSITE; PS00806; ALDOLASE_CLASS_II_2; 1.
SQ SEQUENCE 293 AA; 31316 MW; 3B809F2D20CD0CC CRC64;
Query Match 45.9%; Score 45; DB 2; Length 293;
Best Local Similarity 53.8%; Pred. No. 50;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 4 LKSTAVIPGVV 16
DB 193 LKLTAAVPGFPI 205

RESULT 14

Q7UW43 PRELIMINARY; PRT; 501 AA.
AC Q7UW43;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19).
GN RB2278
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]_taxID=117;
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.C., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Bortym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp. strain 1."
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
DR EMBL; BX294136; CAD72526.1; -.
KW Transferase; Complete proteome.
SQ SEQUENCE 501 AA; 53233 MW; DE4325BF2C49F9BE CRC64;
Query Match 45.9%; Score 45; DB 16; Length 501;
Best Local Similarity 47.4%; Pred. No. 90;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
QY 2 SDLKSTAVIPGVV 20

DB 47 SDAIATVRVPGPVCGSI 65
:::|||||:::
:::|||||:::

RESULT 15

Q94DI0 PRELIMINARY; PRT; 573 AA.
AC Q94DI0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE P0518C01.6 protein.
GN P0518C01.6
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]_taxID=4530;
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC clone: P0518C01."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003277; BAB63669.1; -.
DR Gramene; G94DI0; -.
DR InterPro; IPR002985; PPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF01535; PPR; 7.
DR TrEMBL; TIGR00756; PPR; 7.
SQ SEQUENCE 573 AA; 64109 MW; 4D716B64FC08EFAA CRC64;
Query Match 45.9%; Score 45; DB 10; Length 573;
Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
QY 1 VSDLKSTAVIPGVV 20
DB 381 VFDLXTWTMLIQHCAGEV 400

Search completed: March 10, 2004, 12:10:57
Job time : 33.7917 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:50:51 ; Search time 47.5641 Seconds
(without alignments)
118.807 Million cell updates/sec

Title: US-10-044-703-72

Perfect score: 109

Sequence: 1 YNINISLPSYYPQKSLNY 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Genesep_29Jan04.*

- 1: Genesep1980s.*
- 2: Genesep1990s.*
- 3: Genesep2000s.*
- 4: Genesep2001s.*
- 5: Genesep2002s.*
- 6: Genesep2003as.*
- 7: Genesep2003bs.*
- 8: Genesep2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	20	AAE12287	Aae12287 Mycobacte
2	109	100.0	205	AAp93190	Aap93190 Mycobacte
3	109	100.0	228	AAr71113	Aar71113 Mycobacte
4	109	100.0	228	AAW18161	Aaw18161 Mycobacte
5	109	100.0	228	AAW3036	Aaw3036 Mycobacte
6	109	100.0	228	AAW50742	Aaw50742 Mycobacte
7	109	100.0	228	AAE39353	Aae39353 M. tuberc
8	109	100.0	230	AAW32353	Aaw32353 Mycobacte
9	109	100.0	230	AAW32421	Aaw32421 Mycobacte
10	109	100.0	230	AAW64293	Aaw64293 Mycobacte
11	109	100.0	230	AAW1656	Aaw1656 M. tuberc
12	109	100.0	230	AAW38958	Aaw38958 M. tuberc
13	109	100.0	230	AAW39095	Aaw39095 M. tuberc
14	50	45.9	227	AAW52472	Aaw52472 Mycobacte
15	49.5	45.4	828	ABU31879	Abu31879 Protein e
16	49	45.0	120	AAO10809	Aao10809 Human pol
17	49	45.0	525	ABB70228	Abb70228 Drosophil
18	49	45.0	1805	ABB65262	Abb65262 Drosophil
19	48.5	44.5	631	AAW3105	Aaw3105 C. thermo
20	48	44.0	555	ABB64781	Abb64781 Drosophil
21	48	44.0	576	ABM68877	Abm68877 Phototrab
22	47	43.1	14	AAW74274	Aaw74274 Exemplary
23	47	43.1	14	AAW80880	Aaw80880 Angiostat
24	47	43.1	14	ABG97557	Abg97557 Angiostat
25	47	43.1	309	AAW82257	Aaw82257 S. epider

26	47	43.1	435	6	ABM73565	Abm73565 Staphyloc
27	47	43.1	436	5	ABP38378	Abp38378 Staphyloc
28	47	43.1	1264	2	AAAR36730	Aaar36730 Ubiquitin
29	46	42.2	351	4	AAW52310	Aaw52310 Chicken z
30	46	42.2	454	3	AAW68970	Aaw68970 Cpale pro
31	46	42.2	471	2	AAW61216	Aaw61216 Streptoco
32	46	42.2	471	5	ABP54635	Abp54635 S. pneumo
33	46	42.2	471	7	ADC45239	Adc45239 S. pneumo
34	46	42.2	490	3	AAW81745	Aaw81745 Streptoco
35	46	42.2	506	6	ABU02034	Abu02034 S. pneumo
36	46	42.2	542	4	AAW52304	Aaw52304 Chicken z
37	45	41.3	58	3	AAW03434	Aaw03434 Human sec
38	45	41.3	147	4	AAU20822	Aau20822 Human nov
39	45	41.3	168	3	AAW11898	Aaw11898 Human col
40	45	41.3	168	4	AAW24495	Aaw24495 Colon tum
41	45	41.3	168	6	ABP55344	Abp55344 Human col
42	45	41.3	239	6	ABR39395	Abp39395 Amino aci
43	45	41.3	239	6	ABP98583	Abp98583 Putative
44	45	41.3	275	4	AAW96241	Aaw96241 Putative
45	45	41.3	346	4	ABG20556	Abg20556 Novel hum

ALIGNMENTS

RESULT 1

AAE12287
ID AAE12287 standard; peptide; 20 AA.
XX
AC AAE12287;
XX
DT 18-DEC-2001 (first entry)
XX
DE Mycobacterium tuberculosis (Mtb) peptide #72.
XX
KW Mycobacterium tuberculosis; Mtb peptide; antibacterial; vaccine;
KW infection; anti-Mtb immune response.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200170774-A2.
PD 27-SEP-2001.
XX
PF 20-MAR-2001; 2001WO-US008906.
XX
PR 20-MAR-2000; 2000US-0190834P.
XX
(UYBR-) UNIV BROWN RES FOUND.
PA
PI Degroot AS;
XX
DR WPI; 2001-616401/71.
XX
PT New vaccine for immunizing a mammalian subject, preferably humans,
PT against infection caused by Mycobacterium tuberculosis.
XX
PS Disclosure; fig 4; 42pp; English.
XX
CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine
CC candidate peptides. The invention also relates to a method for
CC identifying Mtb vaccine candidate peptides as well as vaccines comprising
CC these candidate peptides. Vaccines of the invention and Mtb vaccine
CC candidate peptides are useful for inducing an anti-Mycobacterium
CC tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a
CC mammalian subject preferably human. They are used for immunising a
CC mammalian subject, preferably humans, against infection caused by
CC Mycobacterium tuberculosis. The present sequence is a Mtb vaccine
CC candidate peptide
SQ Sequence 20 AA;

Query Match 100.0%; Score 109; DB 4; Length 20;


```

Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YNINISLPSYYPDQKSLNLY 20
DB 1 YNINISLPSYYPDQKSLNLY 20

RESULT 2
AAP93190
ID AAP93190 standard; protein; 205 AA.
XX AAP93190;
XX
XX
XX 26-MAR-1990 (first entry)
XX Mycobacterium bovis BCG-derived MPB64 protein.
XX
XX MPB64 protein; Mycobacterium bovis BCG; vector pKXW64; tuberculosis;
XX amorphous mycobacteriosis.
XX
XX Mycobacterium bovis BCG.
XX
XX JP01247094-A.
XX
XX 02-OCT-1989.
XX
XX 30-MAR-1988; 88JP-00077366.
XX
XX 30-MAR-1988; 88JP-00077366.
XX
XX (AJIN ) AJINOMOTO KK.
XX
XX WPI; 1989-330039/45.
XX N-PSDB; AAN91853.
XX
XX BCG-derived MPB64 protein - used for differentiation of tuberculosis and
XX amorphous mycobacteriosis by antigen-antibody reaction.
XX
XX Claim 1; Page 503; 12pp; Japanese.
XX
XX MPB64 protein is expressed in E.coli using vector pKXW64. It is used for
XX accurate and rapid differentiation of tuberculosis and amorphous
XX mycobacteriosis by antigen-antibody reaction (ELISA method). MPB64
XX protein is produced in large amts. using the vector
XX
XX Sequence 205 AA;

Query Match 100.0%; Score 109; DB 1; Length 205;
Best Local Similarity 100.0%; Pred. No. 3.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YNINISLPSYYPDQKSLNLY 20
DB 27 YNINISLPSYYPDQKSLNLY 46

RESULT 3
AAP71113
ID AAP71113 standard; protein; 228 AA.
XX AAP71113;
XX
XX
XX 25-MAR-2003 (revised)
XX 02-SEP-1995 (first entry)
XX
XX Mycobacterial protein MPT64.
XX
XX Tuberculosis; vaccine; diagnostic.
XX
XX Mycobacterium tuberculosis H37Rv.
XX
XX WO9501440-A1.
PN

```

```

XX 12-JAN-1995.
PD
XX 30-JUN-1994; 94WO-DK000270.
PF
XX 02-JUL-1993; 93DK-00000797.
PR
XX (STAT-) STATENS SERUMINSTITUT.
XX
XX Haslov K, Andersen AB, Oettinger T;
XX WPI; 1995-061004/08.
XX N-PSDB; AAQ84840.
XX
XX Diagnostic skin test for tuberculosis (TB) eliminating positive response
XX in vaccinated animals - also polypeptide(s) and DNA of mycobacterial
XX protein MPT64 and vaccines against TB.
XX
XX Claim 10; Page 51-52; 82pp; English.
XX
XX The polypeptide is an immunological equivalent to MPT64, conferring
XX immunity to tuberculosis caused by bacteria belonging to the tuberculosis
XX complex (Mycobacterium tuberculosis, Mycobacterium africanum and
XX Mycobacterium bovis). Thus, the protein can be used in a vaccine
XX composition against tuberculosis. The vaccine confers greatly increased
XX immunity to tuberculosis, and the method does not result in positive
XX responses in previously vaccinated people. The protein may also be used
XX in a diagnostic kit where it elicits a delayed type hypersensitivity
XX reaction in animals with active tuberculosis, but not in previously
XX vaccinated animals. (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 228 AA;

Query Match 100.0%; Score 109; DB 2; Length 228;
Best Local Similarity 100.0%; Pred. No. 4.3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YNINISLPSYYPDQKSLNLY 20
DB 50 YNINISLPSYYPDQKSLNLY 69

RESULT 4
AAW18161
ID AAW18161 standard; protein; 228 AA.
XX AAW18161;
XX
XX 13-AUG-1997 (first entry)
XX
XX Mycobacterium tuberculosis extracellular 23.5KD protein.
XX
XX Vaccine; vaccinating agent; M.tuberculosis; pathogen; bacteria; virus;
XX fungus; protozoan; HIV.
XX
XX Mycobacterium tuberculosis.
XX
XX Key Location/Qualifiers
XX FT Protein 24..228
XX FT /label= Mature
XX
XX WO9637219-A1.
XX
XX 28-NOV-1996.
XX
XX 23-MAY-1996; 96WO-US0007781.
XX
XX 23-MAY-1995; 95US-00447398.
XX 20-OCT-1995; 95US-00545926.
XX 31-OCT-1995; 95US-00551149.
XX 06-DEC-1995; 95US-00568357.
XX
XX (REGC ) UNIV CALIFORNIA.
PA

```

XX PI Horwitz MA, Harth G;
 XX DR WPI; 1997-020936/02.
 XX DR N-PSDB; AAT71595.
 XX PT Vaccines derived from M.tuberculosis major abundant extracellular
 PT proteins are easy to prepare and less toxic than conventional killed or
 PT attenuated vaccines, useful for protecting against or treating
 PT Mycobacterial infections.
 XX PS Claim 16; Page 40; 193pp; English.
 XX CC A vaccinating agent for promoting an immune response in a mammal against
 CC Mycobacterium pathogens, comprises at least one majorly abundant
 CC extracellular protein, i.e. the M.tuberculosis 110, 80, 71, 58, 45, 32A,
 CC 32B, 30, 24, 23.5, 23, 16, 14 or 12 kD proteins, or their analogues,
 CC homologues and subunits. The present sequence represents the 23.5 kD
 CC protein. The vaccinating agents are used to protect against (or to treat
 CC existing) infections by Mycobacterium (especially M. tuberculosis) while
 CC the epitopes can also be used to detect presence of an immune response to
 CC a Mycobacterium pathogen. The vectors, containing the DNA for the
 CC extracellular proteins, are used to transform cells for production of
 CC recombinant DNA molecules. More generally the DNA from other pathogens
 CC can be used in vaccines, e.g. against other bacteria, viruses, fungi and
 CC protozoa. Since different combinations of DNA can be used, a wide range
 CC of effective compositions can be produced. They generate a response
 CC against the antigens most often found on infected cells during the
 CC infection, regardless of the strength or specificity of the immune
 CC response. The vaccines are easy to produce and less toxic than known
 CC killed or attenuated vaccines, so can be given to immunocompromised
 CC subjects, e.g. those with HIV infection
 XX SQ Sequence 228 AA;
 Query Match 100.0%; Score 109; DB 2; Length 228;
 Best Local Similarity 100.0%; Pred. No. 4.3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YNINISLPSYYPDQKSLNY 20
 Db 50 YNINISLPSYYPDQKSLNY 69
 RESULT 5
 AAW63036
 ID AAW63036 standard; protein; 228 AA.
 AC AAW63036;
 XX 23-OCT-1998 (first entry)
 DE Mycobacterium tuberculosis 23.5 kD protein sequence.
 KW Mycobacterium tuberculosis; vaccination; extracellular product;
 KW immunodominant epitope; interleukin-12; Mf59; immune response;
 KW opsonising humoral response; intracellular pathogen.
 XX OS Mycobacterium tuberculosis.
 XX WO9831388-A1.
 XX 23-JUL-1998.
 XX 15-JAN-1998; 98WO-US000942.
 XX 21-JAN-1997; 97US-00786533.
 XX (REGC) UNIV CALIFORNIA.
 XX Horwitz MA, Harth G, Lee B;
 XX WPI; 1998-413815/35.

DR N-PSDB; AAV42599.
 XX Vaccines against Mycobacterium containing major extracellular proteins -
 PT used to, e.g. induce protective and therapeutic immune responses, and for
 PT detecting an immune response.
 XX PS Example 2; Page 43-45; 236pp; English.
 XX CC This represents a Mycobacterium tuberculosis 23.5 kD protein. The
 CC invention provides an agent for vaccinating mammals against
 CC Mycobacterium. The agent comprises at least one of the major abundant
 CC extracellular 110, 80, 71, 58, 45, 32A, 32B, 30, 24, 23.5, 23, 16, 14 or
 CC 12 kD proteins of M. tuberculosis, or at least 1 of their immunodominant
 CC epitopes and interleukin-12 (IL-12) or Mf59 as adjuvants. The agent
 CC containing the nucleic acid encoding the extracellular products are used
 CC to raise a protective or therapeutic immune response against
 CC Mycobacterium, specifically M. tuberculosis. The immunodominant epitopes
 CC can also be used (typically in a cutaneous hypersensitivity test) to
 CC detect an immune response to vaccination. Preparation of the agent does
 CC not require selection of the most immunogenic products, so large scale
 CC production and purification are easy, resulting in a consistent,
 CC standardised formulation, having lower toxicity than killed or attenuated
 CC vaccines. The agents provide a rapid and effective response (including a
 CC strong cell-mediated component) and are safe even in immunocompromised
 CC subjects. They prevent development of an opsonising humoral response that
 CC might spread intracellular pathogens
 XX SQ Sequence 228 AA;
 Query Match 100.0%; Score 109; DB 2; Length 228;
 Best Local Similarity 100.0%; Pred. No. 4.3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YNINISLPSYYPDQKSLNY 20
 Db 50 YNINISLPSYYPDQKSLNY 69
 RESULT 6
 AAW50742
 ID AAW50742 standard; protein; 228 AA.
 AC AAW50742;
 XX 18-APR-2002 (first entry)
 DE Mycobacterium tuberculosis immunodominant Mtb protein MPT64.
 KW Mtb; MPT64; immunogen; mycobacteria; immunisation; vaccine.
 XX OS Mycobacterium tuberculosis.
 XX Key Location/Qualifiers
 FT Peptide 1..23
 FT /label= Signal peptide
 FT /note= "SEC-dependent signal secretion sequence"
 FT Protein 24..228
 FT /label= Mature_protein
 XX WO200204018-A2.
 XX 17-JAN-2002.
 XX 10-JUL-2001; 2001WO-US021717.
 XX 10-JUL-2000; 2000US-0217646P.
 XX (COLS) UNIV COLORADO STATE RES FOUND.
 XX Orme IM, Belisle JT;
 XX WPI; 2002-164602/21.
 XX

PT Vaccine for boosting immunity to mycobacteria when administered in mid-
 PI life in a subject who has been vaccinated in childhood with Bacillus
 PI Calmette-Guerin, has purified proteins from mycobacterium tuberculosis.
 XX
 PS Claim 8; Page 18; 61pp; English.

XX The present sequence is that of the Mycobacterium tuberculosis (Mtb)
 CC strain H37Rv gene Rv1886c product, designated MPT64. This is one of 31
 CC immunodominant secreted or cytosolic Mtb proteins of strain H37Rv (see
 CC AAM50729-59) discovered through the use of 2-dimensional liquid phase
 CC electrophoresis coupled with an in vitro interferon-gamma assay and
 CC liquid chromatography-mass spectrometry. The immunogens stimulate a
 CC strong interferon-gamma response from T cells of M. tuberculosis infected
 CC mice. The invention provides vaccine compositions for boosting immunity
 CC to mycobacteria when administered in mid-life to a subject who has been
 CC vaccinated neonatally or in early childhood with BCG and in whom
 CC protective immunity has waned. The vaccine compositions comprise 1 or
 CC more of the 31 purified immunogenic proteins. When used as immunogens,
 CC the secreted Mtb proteins lack the secreted signal sequence. A preferred
 CC protein is Ag85A (see AAM50759), the secreted product of the Rv3084v gene
 XX
 SQ Sequence 228 AA;

Query Match 100.0%; Score 109; DB 5; Length 228;
 Best Local Similarity 100.0%; Pred. No. 4.3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YNINISLPSYYPDQKSLNY 20
 |||||
 Db 50 YNINISLPSYYPDQKSLNY 69

RESULT 7

AAE39353
 ID AAE39353 standard; protein; 228 AA.

AC AAE39353;

DT 18-DEC-2003 (first entry)

XX M. tuberculosis extracellular 23.5 KD protein.

XX Vaccine; antibacterial; fungicide; protozoacide; immunostimulant;
 KW virucide; therapy.

OS Mycobacterium tuberculosis.

XX US6599510-B1.

XX 29-JUL-2003.

PF 21-SEP-1998; 98US-00157689.

PR 23-NOV-1993; 93US-00156358.

PR 12-AUG-1994; 94US-00289667.

PR 23-MAY-1995; 95US-00447398.

PR 31-OCT-1995; 95US-00551149.

PR 06-DEC-1995; 95US-00568357.

PR 23-MAY-1996; 96US-00652842.

XX (REG) UNIV CALIFORNIA.

XX Horwitz MA, Harth G;

XX WPI; 2003-669607/63.

DR N-PSDB; AAD59704.

XX A nucleic acid encoding an abundant extracellular protein of
 PT Mycobacterium tuberculosis useful as vaccines for generating protective
 PI or therapeutic immune response against viral, bacterial, fungal or
 PT protozoal infections.
 XX
 PS Example 2; Col 33-34; 82pp; English.

XX

CC The invention relates to a novel nucleic acid encoding an abundant
 CC extracellular protein of Mycobacterium tuberculosis useful as vaccines
 CC for generating protective or therapeutic immune response against viral,
 CC bacterial, fungal and protozoal infections. They are also used as
 CC immunotherapeutic agents. The present sequence is M. tuberculosis strain
 CC Erdman extracellular protein

XX Sequence 228 AA;

Query Match 100.0%; Score 109; DB 7; Length 228;
 Best Local Similarity 100.0%; Pred. No. 4.3e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YNINISLPSYYPDQKSLNY 20
 |||||
 Db 50 YNINISLPSYYPDQKSLNY 69

RESULT 8

AAW32353
 ID AAW32353 standard; protein; 230 AA.

XX AAW32353;

DT 13-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen TbRall.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 KW skin testing; M.tuberculosis.

OS Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT Misc-difference 181

FT /note= "Any amino acid"

XX WO9709429-A2.

XX 13-MAR-1997.

XX 30-AUG-1996; 96WO-US014675.

XX 01-SEP-1995; 95US-00523435.

XX 22-SEP-1995; 95US-00532136.

XX 23-MAR-1996; 96US-00620280.

XX 05-JUN-1996; 96US-00658800.

XX 12-JUL-1996; 96US-00680573.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

XX Vedvick TH, Twardzik DR;

XX WPI; 1997-192904/17.

XX N-PSDB; AAT91451.

XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
 PT useful for diagnosis of M. tuberculosis infection.
 XX
 PS Example 3; Page 101-102; 190pp; English.

XX A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
 CC variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis antigen,
 CC TbRall. The immunogenic polypeptide can be used to diagnose
 CC M.tuberculosis infection by forming complexes with specific antibodies in
 CC the sample. Fragments of DNA encoding the immunogenic polypeptide can be
 CC used as diagnostic primers or probes and agents that bind to the antigen,
 CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
 CC also used for diagnosis

```

XX SQ Sequence 230 AA;
Query Match 100.0%; Score 109; DB 2; Length 230;
Best Local Similarity 100.0%; Pred. No. 4.4e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YNINISLPSYYPDKSLENY 20
Db 116 YNINISLPSYYPDKSLENY 135
|||||
|||||

RESULT 9
AAW32421
ID AAW32421 standard; protein; 230 AA.
XX AC AAW32421;
XX DT 08-JAN-1998 (first entry)
XX DE Mycobacterium tuberculosis antigen Tbrall.
XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX KW skin testing; M.tuberculosis.
XX OS Mycobacterium tuberculosis.
XX FH Key Location/Qualifiers
XX FT Misc-difference 181 /note= "Any amino acid"
XX PN W09709428-A2.
XX PD 13-MAR-1997.
XX PF 30-AUG-1996; 96WO-US014674.
XX PR 01-SEP-1995; 95US-00523436.
XX PR 22-SEP-1995; 95US-00533634.
XX PR 22-MAR-1996; 96US-00620874.
XX PR 05-JUN-1996; 96US-00659683.
XX PR 12-JUL-1996; 96US-00680574.
XX PA (CORI-) CORIXA CORP.
XX PI Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;
XX PI Vedvick TH, Twardzik DR;
XX DR WPI; 1997-132903/17.
XX DR N-PSDB; AAT91515.
XX PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
XX PT useful in vaccines for prevention or treatment of tuberculosis, also for
XX PT diagnosis.
XX PS Example 3; Page 95-96; 168pp; English.
XX CC A new immunogenic polypeptide has been developed comprising an
XX CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
XX CC variant differing only in conservative substitutions and/or
XX CC modifications). The present sequence represents a M.tuberculosis antigen,
XX CC Tbrall. The immunogenic protein, and fusion proteins containing one or
XX CC more of the proteins or one of the proteins plus ESAT-6, are useful in
XX CC vaccines, preferably when formulated with a non-specific adjuvant, to
XX CC induce an immune response against M.tuberculosis (for treatment or
XX CC prevention)
XX SQ Sequence 230 AA;
Query Match 100.0%; Score 109; DB 2; Length 230;
Best Local Similarity 100.0%; Pred. No. 4.4e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 YNINISLPSYYPDKSLENY 20
Db 116 YNINISLPSYYPDKSLENY 135
|||||
|||||

RESULT 10
AAW64293
ID AAW64293 standard; protein; 230 AA.
XX AC AAW64293;
XX DT 17-OCT-2003 (revised)
XX DT 09-NOV-1998 (first entry)
XX DE Mycobacterium tuberculosis antigen Tbrall.
XX KW Tuberculosis; infection; diagnosis; antigen; Tbrall.
XX OS Mycobacterium tuberculosis; strain H37Ra.
XX FH Key Location/Qualifiers
XX FT Misc-difference 181 /note= "encoded by AMG"
XX PN W09816645-A2.
XX PD 23-APR-1998.
XX PF 07-OCT-1997; 97WO-US018214.
XX PR 11-OCT-1996; 96US-00729622.
XX PR 13-MAR-1997; 97US-00818111.
XX PA (CORI-) CORIXA CORP.
XX PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX PI Vedvick TS, Twardzik DR, Lodes MJ;
XX DR WPI; 1998-251292/22.
XX DR N-PSDB; AAV44341.
XX PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
XX PT develop products for the detection of M. tuberculosis infection and
XX PT diagnosis of tuberculosis.
XX PS Example 3; Page 99-100; 250pp; English.
XX CC This polypeptide comprises Mycobacterium tuberculosis soluble antigen
XX CC Tbrall. It is encoded by a DNA sequence (see AAV44341) isolated from a M.
XX CC tuberculosis strain H37Ra expression library with rabbit anti-aera raised
XX CC against M. tuberculosis supernatant. Tbrall has previously been
XX CC identified in M. tuberculosis. The invention relates to compositions and
XX CC methods for diagnosing tuberculosis. It provides polypeptides (see
XX CC AAW64291-W64379) comprising an antigenic portion of a soluble M.
XX CC tuberculosis antigen, or an immunogenic portion of an M. tuberculosis
XX CC antigen, as well as DNA sequences encoding such polypeptides, recombinant
XX CC expression vectors and transformed or transfected host cells. Also
XX CC claimed are methods and diagnostic kits for detecting M. tuberculosis
XX CC infection in a patient using the above polypeptides, antibodies, or
XX CC oligonucleotide probes and primers, for the diagnosis of tuberculosis.
XX CC (Updated on 17-OCT-2003 to standardise OS field)
XX SQ Sequence 230 AA;
Query Match 100.0%; Score 109; DB 2; Length 230;
Best Local Similarity 100.0%; Pred. No. 4.4e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 YNINISLPSYYPDKSLENY 20
Db 116 YNINISLPSYYPDKSLENY 135
|||||
|||||

```

RESULT 11

AAW81656
ID AAW81656 standard; protein; 230 AA.
XX
AC AAW81656;
XX
DT 27-JAN-1999 (first entry)
XX
DE M. tuberculosis immunogenic polypeptide Tbrall.
XX
XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT Misc-difference 181
FT /label= unknown
XX
PN W09816646-A2.
XX
PD 23-APR-1998.
XX
PF 07-OCT-1997; 97WO-US018293.
XX
PR 11-OCT-1996; 96US-00730510.
PR 13-MAR-1997; 97US-00818112.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ;
XX
DR WPI; 1998-261042/23.
DR N-PSDB; AAW64449.
XX
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
PT develop products for the detection of M. tuberculosis infection and for
PT diagnosis; treatment and prevention of tuberculosis.
XX
PS Example 3; Page 96-97; 230pp; English.

XX
CC This sequence represents an immunogenic portion of a soluble
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
CC inducing protective immunity against tuberculosis (TB). This sequence can
CC be formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis
XX
SQ Sequence 230 AA;
Query Match 100.0%; Score 109; DB 2; Length 230;
Best Local Similarity 100.0%; Pred. No. 4.4e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YNINISLPSYYPDQKSLNLY 20
DB 116 YNINISLPSYYPDQKSLNLY 135

RESULT 12

AAW81656
ID AAW81656 standard; protein; 230 AA.
XX
AC AAW81656;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis recombinant antigen protein Tbrall.
XX
XX Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity.
XX

OS Mycobacterium tuberculosis.

XX
PN W09942118-A2.
XX
PD 26-AUG-1999.
XX
PF 17-FEB-1999; 99WO-US003265.
XX
PR 18-FEB-1998; 98US-00024753.
PR 05-MAY-1998; 98US-00072596.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX
DR WPI; 1999-527416/44.
DR N-PSDB; AA219039.
XX
PT New polypeptide comprising antigenic portions of M. tuberculosis.
PS Example 3; Page 137; 323pp; English.

XX
CC This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against M.
CC tuberculosis infection. The new detection methods are needed as current
CC vaccination strategies do not provide 100% immunity
XX
SQ Sequence 230 AA;

Query Match 100.0%; Score 109; DB 2; Length 230;
Best Local Similarity 100.0%; Pred. No. 4.4e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YNINISLPSYYPDQKSLNLY 20

DB 116 YNINISLPSYYPDQKSLNLY 135

RESULT 13

AAW81656
ID AAW81656 standard; protein; 230 AA.
XX
AC AAW81656;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis antigen Tbrall amino acid sequence.
XX
XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
OS Mycobacterium tuberculosis.
XX
PN W09942076-A2.
XX
PD 26-AUG-1999.
XX
PF 17-FEB-1999; 99WO-US003268.
XX
PR 18-FEB-1998; 98US-00025197.
PR 05-MAY-1998; 98US-00072967.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX
DR WPI; 1999-527409/44.

DR N-PSDB; AAZ19251.
 XX New antigens from Mycobacterium tuberculosis useful in diagnostic skin
 PT tests and protective or therapeutic vaccines or compositions.
 XX
 XX Example 3; Page 97; 299pp; English.
 XX
 XX The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
 CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
 CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
 CC polypeptides fragments, can be used in pharmaceutical compositions or
 CC vaccines to generate a protective or therapeutic immune response to M.
 CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
 CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
 CC killer cells and/or macrophages in tuberculosis-immune subjects. AA219249
 CC to AA219460 and AA39083 to AA39225 are used in the exemplification of
 CC the present invention
 XX
 XX Sequence 230 AA;
 SQ
 Query Match 100.0%; Score 109; DB 2; Length 230;
 Best Local Similarity 100.0%; Pred. No. 4.4e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YNINISLPSPYDPQKSLNLY 20
 DB 116 YNINISLPSPYDPQKSLNLY 135
 RESULT 14
 AAB52472
 ID AAB52472 standard; protein; 227 AA.
 XX
 XX AAB52472;
 XX
 XX 23-FEB-2001 (first entry)
 DT
 XX
 XX Mycobacterium tuberculosis secreted protein #37.
 DE
 XX
 XX Mycobacterium tuberculosis secreted protein; MTSP; vaccine.
 KW
 XX
 XX Mycobacterium tuberculosis.
 OS
 XX
 XX WO200066143-A1.
 PN
 XX
 XX 09-NOV-2000.
 PD
 XX
 XX 04-MAY-2000; 2000WO-US012197.
 PF
 XX
 XX 04-MAY-1999; 99US-0132479P.
 PR
 XX 04-MAY-1999; 99US-0132503P.
 PR
 XX (PUBL-) PUBLIC HEALTH RES INST NEW YORK.
 FA
 XX
 XX Gennaro ML, Gomez MJ;
 FI
 XX
 XX WPI; 2001-007151/01.
 DR
 XX
 XX Novel Mycobacterium tuberculosis secreted polypeptides and
 PT polynucleotides useful in diagnosis, treatment and prophylaxis of
 PT tuberculosis.
 PT
 XX
 XX Claim 11; Fig 1; 60pp; English.
 PS
 XX
 XX The present invention relates to Mycobacterium tuberculosis secreted
 CC proteins (MTSP), where the polypeptide has M. tuberculosis specific
 CC antigenic and immunogenic properties. Compositions of the invention may
 CC be useful for diagnosing Mycobacterium tuberculosis infection and as a
 CC vaccine against M. tuberculosis infection
 CC
 XX Sequence 227 AA;
 SQ

Query Match 45.9%; Score 50; DB 4; Length 227;
 Best Local Similarity 40.0%; Pred. No. 11;
 Matches 8; Conservative 7; Mismatches 5; Indels 0; Gaps 0;
 QY 1 YNINISLPSPYDPQKSLNLY 20
 DB 49 YMLDWTFRVDYDQQAALTDY 68
 RESULT 15
 ABU31879
 ID ABU31879 standard; protein; 828 AA.
 XX
 XX AC ABU31879;
 XX
 XX DT 19-JUN-2003 (first entry)
 XX
 XX DE Protein encoded by Prokaryotic essential gene #17406.
 XX
 XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 XX OS Klebsiella pneumoniae.
 XX
 XX PN WO200277183-A2.
 XX
 XX PD 03-OCT-2002.
 XX
 XX PF 21-MAR-2002; 2002WO-US009107.
 XX
 XX PR 21-MAR-2001; 2001US-00815242.
 PR
 XX 06-SEP-2001; 2001US-00948993.
 PR
 XX 25-OCT-2001; 2001US-0342923P.
 PR
 XX 08-FEB-2002; 2002US-00072851.
 PR
 XX 06-MAR-2002; 2002US-0362699P.
 XX
 XX PA (ELIT-) ELITRA PHARM INC.
 XX
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 XX WPI; 2003-029926/02.
 DR
 XX N-PSDB; ACA35749.
 XX
 XX PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 PT
 XX
 XX PS Claim 25; SEQ ID NO 59803; 1766pp; English.
 XX
 XX CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 828 AA;

Query Match 45.4%; Score 49.5; DB 6; Length 828;
 Best Local Similarity 55.6%; Pred. No. 63;
 Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 3 INISLPYYPDOKSLENY 20
 :|||:| | ||:| ||
 Db 144 LNISVPQIYEDQR-LKNY 160

Search completed: March 10, 2004, 12:05:36
 Job time : 49.5641 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 12:11:07 ; Search time 25:1282 Seconds
(without alignments)
168.061 Million cell updates/sec

Title: US-10-044-703-72

Perfect score: 109
Sequence: 1 YNINISLPSYDPQKSLNY 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	100.0	20	9 US-09-813-333-72	Sequence 72, Appl
2	109	100.0	20	13 US-10-044-703-72	Sequence 72, Appl
3	109	100.0	230	14 US-10-193-002-66	Sequence 66, Appl
4	109	100.0	230	14 US-10-084-843-65	Sequence 65, Appl
5	86	78.9	21	14 US-10-082-710-180	Sequence 180, App
6	67	61.5	21	14 US-10-082-710-179	Sequence 179, App
7	47	43.1	14	9 US-09-766-412-48	Sequence 48, Appl
8	47	43.1	1272	15 US-10-369-493-22374	Sequence 22374, A
9	46	42.2	351	14 US-10-239-431A-22	Sequence 22, Appl
10	46	42.2	454	9 US-09-767-041-30	Sequence 30, Appl
11	46	42.2	471	9 US-09-765-272-158	Sequence 158, App
12	46	42.2	490	10 US-09-749-744A-98	Sequence 98, Appl
13	46	42.2	542	14 US-10-239-431A-10	Sequence 10, Appl
14	45	41.3	168	9 US-09-922-217-198	Sequence 198, App
15	45	41.3	168	9 US-09-833-263-198	Sequence 198, App

16	45	41.3	168	13 US-10-025-380-198	Sequence 198, App
17	45	41.3	225	15 US-10-369-493-19679	Sequence 19679, A
18	45	41.3	239	15 US-10-189-507-6	Sequence 6, Appl
19	45	41.3	239	15 US-10-189-507-10	Sequence 10, Appl
20	45	41.3	260	15 US-10-369-493-5542	Sequence 5542, Ap
21	45	41.3	275	15 US-10-369-493-1272	Sequence 1272, Ap
22	45	41.3	275	15 US-10-369-493-20338	Sequence 20338, A
23	45	41.3	275	15 US-10-369-493-21609	Sequence 21609, A
24	45	41.3	478	13 US-10-193-295-2	Sequence 2, Appl
25	45	41.3	478	13 US-10-622-516-2	Sequence 2, Appl
26	45	41.3	518	13 US-10-193-295-5	Sequence 5, Appl
27	45	41.3	518	13 US-10-622-516-5	Sequence 5, Appl
28	45	41.3	520	13 US-10-193-295-4	Sequence 4, Appl
29	45	41.3	520	14 US-10-354-358-56	Sequence 56, Appl
30	45	41.3	520	15 US-10-622-516-4	Sequence 2, Appl
31	45	41.3	575	9 US-09-735-932-4	Sequence 2, Appl
32	45	41.3	575	9 US-09-735-932-4	Sequence 1, Appl
33	45	41.3	575	9 US-09-927-267-1	Sequence 16, Appl
34	45	41.3	575	9 US-09-927-267-16	Sequence 30, Appl
35	45	41.3	575	10 US-09-842-758-30	Sequence 74, Appl
36	45	41.3	575	10 US-09-842-758-74	Sequence 28, Appl
37	45	41.3	575	14 US-10-207-951-2	Sequence 7931, Ap
38	45	41.3	578	10 US-09-842-758-28	Sequence 4, Appl
39	45	41.3	1077	14 US-10-032-585-7931	Sequence 2983, Ap
40	45	41.3	1704	14 US-10-207-951-4	Sequence 44, Appl
41	44	40.4	397	15 US-10-094-749-2983	Sequence 44, Appl
42	44	40.4	430	9 US-09-861-695-44	Sequence 44, Appl
43	44	40.4	430	9 US-09-464-099A-44	Sequence 44, Appl
44	44	40.4	456	10 US-09-932-227-64	Sequence 64, Appl
45	44	40.4	456	14 US-10-183-708-64	

ALIGNMENTS

RESULT 1
US-09-813-333-72
; Sequence 72, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US/09/813,333
; CURRENT APPLICATION NUMBER: US/09/813,333
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 72
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-72

Query Match 100.0%; Score 109; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.2e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YNINISLPSYDPQKSLNY 20
Db 1 YNINISLPSYDPQKSLNY 20

RESULT 2
US-10-044-703-72
; Sequence 72, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703

;; CURRENT FILING DATE: 2002-05-20
;; PRIOR APPLICATION NUMBER: 60/190,834
;; PRIOR FILING DATE: 2000-03-20
;; NUMBER OF SEQ ID NOS: 81
;; SOFTWARE: Patent In Ver. 2.1
;; SEQ ID NO 72
;; LENGTH: 20
;; TYPE: PRT
;; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-72

Query Match 100.0%; Score 109; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.2e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YNINISLPSYYPDKSLNY 20
|||
Db 1 YNINISLPSYYPDKSLNY 20
|||

RESULT 3

US-10-193-002-66
; Sequence 66, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

;; Skeiky, Yasir A.W.
;; Dillon, Davin C.
;; Campos-Neto, Antonia
;; Houghton, Raymond
;; Vedvick, Thomas S.
;; Twardzik, Daniel R.
;; Lodes, Michael J.
;; Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 230 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 66:

US-10-193-002-66

Query Match

100.0%; Score 109; DB 14; Length 230;

Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 YNINISLPSYYPDKSLNY 20
|||
Db 116 YNINISLPSYYPDKSLNY 135
|||

RESULT 4

US-10-084-843-65
; Sequence 65, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

;; Skeiky, Yasir A.W.
;; Dillon, Davin C.
;; Campos-Neto, Antonio
;; Houghton, Raymond
;; Vedvick, Thomas S.
;; Twardzik, Daniel R.
;; Lodes, Michael J.
;; Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 230 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 65:

US-10-084-843-65

Query Match 100.0%; Score 109; DB 14; Length 230;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YNINISLPSYYPDKSLNY 20
|||
Db 116 YNINISLPSYYPDKSLNY 135
|||

RESULT 5

US-10-062-710-180
; Sequence 180, Application US/10062710
; Publication No. US20030049253A1

GENERAL INFORMATION:
; APPLICANT: Li, Frank Q.
; APPLICANT: Chu, Yong-Liang
; APPLICANT: Qiu, Jian-Tai
; TITLE OF INVENTION: Polymorphic Conjugates for Delivery of
; TITLE OF INVENTION: MHC-Recognized Epitopes
; TITLE OF INVENTION: Via Peptide Vaccines
; FILE REFERENCE: 3781-001-27
; CURRENT APPLICATION NUMBER: US/10/062,710
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/310,498
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: MPT64 protein
US-10-062-710-180

Query Match 78.9%; Score 86; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ISLPSYYPDQKSLNY 20
| | | | | | | | | | | | | | | | | | | | | |
Db 2 ISLPSYYPDQKSLNY 17

RESULT 6
US-10-062-710-179
; Sequence 179, Application US/10062710
; Publication No. US20030049253A1
; GENERAL INFORMATION:
; APPLICANT: Li, Frank Q.
; APPLICANT: Chu, Yong-Liang
; APPLICANT: Qiu, Jian-Tai
; TITLE OF INVENTION: Polymorphic Conjugates for Delivery of
; TITLE OF INVENTION: MHC-Recognized Epitopes
; TITLE OF INVENTION: Via Peptide Vaccines
; FILE REFERENCE: 3781-001-27
; CURRENT APPLICATION NUMBER: US/10/062,710
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/310,498
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 179
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: MPT64 protein
US-10-062-710-179

Query Match 61.5%; Score 67; DB 14; Length 21;
Best Local Similarity 92.9%; Pred. No. 0.002;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YNINISLPSYYPDQ 14
| | | | | | | | | | | | | | | | | | | | | |
Db 8 YNINISLPSYYPDQ 21

RESULT 7
US-09-766-412-48
; Sequence 48, Application US/09766412
; Patent No. US20020103129A1
; GENERAL INFORMATION:
; APPLICANT: GE, Ruowen et al.
; TITLE OF INVENTION: SMALL PEPTIDES HAVING ANTI-ANGIOGENIC AND ENDOTHELIAL CELL INHIBIT

GENERAL INFORMATION:
; FILE REFERENCE: 1781-0215P
; CURRENT APPLICATION NUMBER: US/09/766,412
; CURRENT FILING DATE: 2001-01-11
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Mammalian
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: h/mFLK2
US-09-766-412-48

Query Match 43.1%; Score 47; DB 9; Length 14;
Best Local Similarity 58.3%; Pred. No. 1.4;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 2 NINISLPSYYPD 13
| | | | | | | | | | | | | | | | | | | | | |
Db 3 NINISLPSYYPD 14

RESULT 8
US-10-369-493-22374
; Sequence 22374, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22374
; LENGTH: 1272
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22374

Query Match 43.1%; Score 47; DB 15; Length 1272;
Best Local Similarity 44.4%; Pred. No. 2.1e+02;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 3 INISLPSYYPDQKSLNY 20
| | | | | | | | | | | | | | | | | | | | | |
Db 449 INISLPSYYPDQKSLNY 466

RESULT 9
US-10-239-431A-22
; Sequence 22, Application US/10239431A
; Publication No. US20030170726A1
; GENERAL INFORMATION:
; APPLICANT: FRADELIZE, JULIE
; APPLICANT: FRIEDERICH, EVELYNE
; APPLICANT: GOLSTEIN, ROY W.
; APPLICANT: LOUWARD, DANIEL
; APPLICANT: NOIREAUX, VINCENT
; APPLICANT: SYKES, CECILE
; TITLE OF INVENTION: PEPTIDE SEQUENCES COMPRISING ONE OR MORE UNITS BINDING
; TITLE OF INVENTION: TO PROTEINS OF THE Ena/VASP FAMILY, AND THEIR USES
; FILE REFERENCE: 0508-1032
; CURRENT APPLICATION NUMBER: US/10/239,431A
; CURRENT FILING DATE: 2002-09-23

;; PRIOR APPLICATION NUMBER: PCT/FR01/00843
;; PRIOR FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: FR 00/03637
;; PRIOR FILING DATE: 2000-03-22
;; NUMBER OF SEQ ID NOS: 38
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 22
;; LENGTH: 351
;; TYPE: PRT
;; ORGANISM: Gallus gallus
US-10-239-431A-22

Query Match 42.2%; Score 46; DB 14; Length 351;
Best Local Similarity 64.3%; Pred. No. 70;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 NINISLPYYPDK 15
Db 16 SINISTPFPNPK 29

RESULT 10
US-09-767-041-30
;; Sequence 30, Application US/09767041
;; Patent No. US2002005168A1
;; GENERAL INFORMATION:
;; APPLICANT: Smith, Hilda
;; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS
;; FILE REFERENCE: 2183-4726
;; CURRENT APPLICATION NUMBER: US/09/767,041
;; CURRENT FILING DATE: 2001-01-22
;; PRIOR APPLICATION NUMBER: PCT/NL99/00460
;; PRIOR FILING DATE: 1999-07-19
;; PRIOR APPLICATION NUMBER: EP98202465.5
;; PRIOR FILING DATE: 1998-07-22
;; PRIOR APPLICATION NUMBER: EP98202467.1
;; PRIOR FILING DATE: 1998-07-22
;; NUMBER OF SEQ ID NOS: 53
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 30
;; LENGTH: 454
;; TYPE: PRT
;; ORGANISM: Streptococcus suis
;; FEATURE:
;; NAME/KEY: misc feature
;; OTHER INFORMATION: CP51E
US-09-767-041-30

Query Match 42.2%; Score 46; DB 9; Length 454;
Best Local Similarity 50.0%; Pred. No. 94;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 3 INISLPYYPDKSLE 18
Db 168 INLSPLYYSVEATE 193

RESULT 11
US-09-765-272-158
;; Sequence 158, Application US/09765272
;; Patent No. US20020061545A1
;; GENERAL INFORMATION:
;; APPLICANT: Choi et. al.
;; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
;; NUMBER OF SEQUENCES: 452
;; CORRESPONDENCE ADDRESS:
;; ADDRESSER: Human Genome Sciences, Inc.
;; STREET: 9410 Key West Avenue
;; CITY: Rockville
;; STATE: Maryland
;; COUNTRY: USA
;; ZIP: 20850
;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;; COMPUTER: HP Vectra 486/33
;; OPERATING SYSTEM: MSDOS version 6.2
;; SOFTWARE: ASCII Text
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/765,272
;; FILING DATE: 22-Jan-2001
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/961,083
;; FILING DATE: <Unknown>
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Brookes, A. Anders
;; REGISTRATION NUMBER: 36,373
;; REFERENCE/DOCKET NUMBER: PB340P2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (301) 309-8504
;; TELEFAX: (301) 309-8512
;; INFORMATION FOR SEQ ID NO: 158:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 471 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; SEQUENCE DESCRIPTION: SEQ ID NO: 158:
US-09-765-272-158

Query Match 42.2%; Score 46; DB 9; Length 471;
Best Local Similarity 45.0%; Pred. No. 98;
Matches 9; Conservative 4; Mismatches 3; Indels 4; Gaps 1;

Qy 1 YNINISLPYYPDKSLENY 20
Db 355 YNNNLSYPIYY----DVENV 370

RESULT 12
US-09-769-744A-98
;; Sequence 98, Application US/09769744A
;; Publication No. US20030134407A1
;; GENERAL INFORMATION:
;; APPLICANT: Le Page, Richard WF
;; APPLICANT: Wells, Jeremy M
;; APPLICANT: Hanniffy, Sean B
;; APPLICANT: Hanebro, Philip M
;; TITLE OF INVENTION: Proteins
;; FILE REFERENCE: PWC/F21122WO
;; CURRENT APPLICATION NUMBER: US/09/769,744A
;; CURRENT FILING DATE: 2001-01-26
;; PRIOR APPLICATION NUMBER: PCT/GB99/02452
;; PRIOR FILING DATE: 1999-07-27
;; PRIOR APPLICATION NUMBER: GB 9816336.3
;; PRIOR FILING DATE: 1998-07-27
;; PRIOR APPLICATION NUMBER: US 60/125329
;; PRIOR FILING DATE: 1999-03-19
;; NUMBER OF SEQ ID NOS: 196
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 98
;; LENGTH: 490
;; TYPE: PRT
;; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-98

Query Match 42.2%; Score 46; DB 10; Length 490;
Best Local Similarity 45.0%; Pred. No. 1e+02;
Matches 9; Conservative 4; Mismatches 3; Indels 4; Gaps 1;

Qy 1 YNINISLPYYPDKSLENY 20
Db 374 YNNNLSYPIYY----DVENV 389

```
RESULT 13
US-10-239-431A-10
; Sequence 10, Application US/10239431A
; Publication No. US20030170726A1
; GENERAL INFORMATION:
; APPLICANT: FRADELIZE, JULIE
; APPLICANT: FRIEDERICH, EVELYNE
; APPLICANT: GOLSTEYN, ROY M.
; APPLICANT: LOUVARD, DANIEL
; APPLICANT: NOIREAUX, VINCENT
; APPLICANT: SYKES, CECILE
; TITLE OF INVENTION: PEPTIDE SEQUENCES COMPRISING ONE OR MORE UNITS BINDING
; TO PROTEINS OF THE Bna/VASP FAMILY, AND THEIR USES
; FILE REFERENCE: 0508-1032
; CURRENT APPLICATION NUMBER: US/10/239.431A
; CURRENT FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: PCT/FR01/00843
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: FR 00/03637
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 10
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Gallus gallus
US-10-239-431A-10
Query Match 42.2%; Score 46; DB 14; Length 542;
Best Local Similarity 64.3%; Pred. No. 1.1e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 NINISLPSYYPDQX 15
: |||||:|:|
DB 16 SINISTPSFYNPQX 29

RESULT 14
US-09-922-217-198
; Sequence 198, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yudiu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922.217
; CURRENT FILING DATE: 2001-09-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 198
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-198
Query Match 41.3%; Score 45; DB 9; Length 168;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 3 INISLPSYYPDQKSLENY 20
: |||||:|:|
```

```
DB 60 LEIYFPSQYVDQAELEKY 77

RESULT 15
US-09-833-263-198
; Sequence 198, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833.263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 198
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-833-263-198
Query Match 41.3%; Score 45; DB 9; Length 168;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 3 INISLPSYYPDQKSLENY 20
: |||||:|:|
DB 60 LEIYFPSQYVDQAELEKY 77

Search completed: March 10, 2004, 12:41:32
Job time : 26.1282 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:58:01 ; Search time 10.7692 Seconds
(without alignments)
178.641 Million cell updates/sec

Title: US-10-044-703-72

Perfect score: 109

Sequence: 1 YNINISLPSPYDPQKSLNLY 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:**

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	109	100.0	228	2	B70756	mpt64 protein - My
2	109	100.0	228	2	A30545	immunogenic protei
3	50	45.9	227	2	G70859	probable secreted
4	49	45.0	675	2	T33560	hypothetical prote
5	49	45.0	1433	1	A36734	bacillopeptidase F
6	48.5	44.5	631	2	JC6031	scaffoldin dockeri
7	47	43.1	429	2	F63063	tryptophan synthas
8	47	43.1	432	2	H89924	3-phosphoshikimate
9	47	43.1	1272	2	S60999	ubiquitin-specific
10	46	42.2	354	2	T27112	hypothetical prote
11	46	42.2	501	2	F98050	lysozyme (EC 3.2.1
12	46	42.2	542	2	A44358	zykin - chicken
13	46	42.2	1007	2	H81670	conserved hypothet
14	45.5	41.7	287	2	G81699	peptide ABC transp
15	45	41.3	260	2	T28119	hypothetical prote
16	45	41.3	275	2	G75130	translation initia
17	45	41.3	275	2	D71087	probable translati
18	45	41.3	520	2	S45497	hydroxymethylgluta
19	45	41.3	520	2	S27197	hydroxymethylgluta
20	45	41.3	520	2	A25332	hydroxymethylgluta
21	45	41.3	520	2	S12736	hydroxymethylgluta
22	45	41.3	522	2	S13887	hydroxymethylgluta
23	45	41.3	575	2	I53927	olfactory cyclic n
24	45	41.3	1135	2	T42368	DNA-directed RNA p
25	44.5	40.8	148	2	S78271	hypothetical prote
26	44.5	40.8	187	2	C75136	hypothetical prote
27	44.5	40.8	532	2	B82354	deoxycytidylate de
28	44	40.4	128	2	G82594	hypothetical prote
29	44	40.4	239	2	D90470	hypothetical prote

30 44 40.4 270 2 D84216 cytochrome b6 [imp
31 44 40.4 343 2 F90258 conserved hypothet
32 44 40.4 455 2 A53565 hydroxymethylgluta
33 44 40.4 1037 2 A36096 Ca2+-transporting
34 44 40.4 1135 2 T42723 probable DNA-dirc
35 43.5 39.9 348 2 F90264 acetoin utilization
36 43.5 39.9 919 1 RNZCA DNA-directed RNA p
37 43.5 39.9 1165 2 T30731 probable DNA-dirc
38 43 39.4 64 2 S74342 transposase ssi076
39 43 39.4 111 2 S70302 hypothetical prote
40 43 39.4 136 2 AG1951 hypothetical prote
41 43 39.4 153 2 S48998 hypothetical prote
42 43 39.4 332 2 T21279 hypothetical prote
43 43 39.4 336 2 D86710 collagen adhesin
44 43 39.4 343 2 AB2045 hypothetical prote
45 43 39.4 370 2 JE0342 E2F transcription

ALIGNMENTS

RESULT 1

B70756
mpt64 protein - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C/Accession: B70756; S38198
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, R.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, J.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: B70756
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-228 <COL>
A/Cross-references: GB:274025; GB:AL123456; NID:G3261586; PIDN:CAA98382.1; PID:G140347;
A/Experimental source: strain H37RV
R/Oettinger, T.; Andersen, A.B.
submitted to the EMBL Data Library, October 1993
A/Description: Cloning and B-cell epitope mapping of MPT64 from Mycobacterium tuberculosis
A/Reference number: S38198
A/Accession: S38198
A/Molecule type: mRNA
A/Residues: 1-228 <OET>
A/Cross-references: EMBL:X75361; NID:G407149; PIDN:CAA53143.1; PID:G581375
R/Nagai, S.; Wiker, H.G.; Harboe, M.; Kinomoto, M.
Infect. Immun. 59, 372-382, 1991
A/Title: Isolation and partial characterization of major protein antigens in the cultu
A/Reference number: A60274; MUID:91099989; PMID:1898899
C/Contents: annotation; amino terminal sequence
C/Genetics:
A/Gene: mpt64
A/Start codon: GTG
F/2-228/Product: mpt64 protein #status predicted <MAT>

Query Match 100.0%; Score 109; DB 2; Length 228;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YNINISLPSPYDPQKSLNLY 20
|||
DB 50 YNINISLPSPYDPQKSLNLY 69

RESULT 2

A30545
immunogenic protein MP564 precursor - Mycobacterium bovis
N/Alternate names: antigen MP564
C/Species: Mycobacterium bovis
C/Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 15-Jun-1996

```

Query Match          45.9%; Score 50; DB 2; Length 227;
Best Local Similarity 40.0%; Pred.No. 2.9;
Matches 8; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY      1 YNINISLPSVYPDOKSLENY 20
          | : : : : | : : : : |
Db      49 YMLDMTFVDPDQCALTDY 68

RESULT 4
T33560
hypothetical protein F34D6.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33560
R:Bauer, C.; Wohldmann, P.; Courtney, L.
submitted to the EMBL Data Library, Octob 1998
A:Description: The sequence of C. elegans cosmid Y47G7B.
A:Reference number: Z21369
A:Accession: T33560
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-675 <BAU>
A:Cross-references: EMBL:AF099000; PIDN:AAC68732.1; GSPDB:GN00020; CESP:F34D6.5
A:Experimental source: strain Bristol N2; clone Y47G7B
A:Genetics:
A:Gene: CESP:F34D6.5
A:Map position: 94
A:Introns: 52/3; 94/3; 167/3; 275/1; 314/2; 331/3; 393/3; 435/3; 508/3; 613/1; 652/2

```

A/Cross-references: GB:M22630; NID:g142938; PIDN:AAA23458.1; PID:g551705
A/Resumes: 1-211 KRES3
R/Kunst, F.; Ogasawara, N.; Moszer, I.; Alberini, A.M.; Alloni, G.; Azevedo, V.; Berter

C;Accession: S60999; A44450; S61680; S67009; S63876; S27466
R;Wiemann, S.; Rechmann, S.; Benes, V.; Voss, H.; Schwager, C.; Vlcek, C.; Stegemann, J.
submitted to the EMBL Data Library, August 1995
A;Description: Sequencing of 51 kilobases on the right arm of chromosome XV from *S. cere*
A;Reference number: S60998
A;Accession: S60999
A;Molecule type: DNA
A;Residues: 1-1272 <WIE>
A;Cross-references: EMBL:X90518; NID:G1050808; PID:G1050825
J;Baker, R.T.; Tobias, J.W.; Varshavsky, A.
J. Biol. Chem. 267, 23364-23375, 1992
A;Title: Ubiquitin-specific proteases of *Saccharomyces cerevisiae*. Cloning of USP2 and U
A;Reference number: A44450; MUID:93054674; PMID:1429680
A;Accession: A44450
A;Molecule type: DNA
A;Residues: 1-655,664-1272 <BAK>
A;Cross-references: EMBL:M94916; NID:G173127; PID:AAA35190.1; PID:G173128
A;Note: sequence extracted from NCBI backbone (NCBI:118306)
R;Benes, V.; Andrade, M.A.; Rechmann, S.; Teodoru, C.; Banrevi, A.; Sander, C.; Valencia
submitted to the EMBL Data Library, December 1995
A;Description: Nucleotide sequence and analysis of a 130 kb fragment of yeast chromosome
A;Reference number: S61680
A;Accession: S61680
A;Molecule type: DNA
A;Residues: 1-1272 <BEN>
A;Cross-references: EMBL:X94335; NID:G1262139; PID:G1164968
R;Voss, H.; Benes, V.; Rechmann, S.; Teodoru, C.; Schwager, C.; Paces, V.; Ansoorge, W.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S66965
A;Accession: S67009
A;Molecule type: DNA
A;Residues: 1-1272 <VOS>
A;Cross-references: EMBL:Z75032; NID:G1420323; PID:G1420324; MIPS:YOR124C
A;Experimental source: strain S288C
R;Wiemann, S.; Rechmann, S.; Benes, V.; Voss, H.; Schwager, C.; Vlcek, C.; Stegemann, J.
Yeast 12, 281-288, 1996
A;Title: Sequencing and analysis of 51 kb on the right arm of chromosome XV from *Sacchar*
A;Reference number: S63876
A;Accession: S63876
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1272 <WIV>
A;Cross-references: EMBL:X90518; NID:G1050808; PID:G1050825
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
C;Genetics:
A;Gene: SGD:UBP2
A;Cross-references: SGD:S0005650; MIPS:YOR124C
A;Map position: 15R
C;Keywords: hydrolase

Query Match 43.1%; Score 47; DB 2; Length 1272;
Best Local Similarity 44.4%; Pred. No. 65;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 3 INISLPSVYPDQKSLNLY 20
Db 449 INLSVSHYTRDILKNY 466

RESULT 10
T27712
hypothetical protein ZK1225.1 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
C;Accession: T27712
R;Kershaw, J.
submitted to the EMBL Data Library, March 1998
A;Reference number: Z20410
A;Accession: T27712
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-354 <WIL>
A;Cross-references: EMBL:AL022289; PIDN:CAA18372.1; GSPDB:GN00019; CBSP:ZK1225.1

A;Experimental source: clone ZK1225
C;Genetics:
A;Gene: CBSP:ZK1225.1
A;Map position: 1
A;Introns: 4/2; 45/3; 91/1; 248/1; 286/2; 321/2
C;Superfamily: *Caenorhabditis elegans* hypothetical protein P2285.1

Query Match 42.2%; Score 46; DB 2; Length 354;
Best Local Similarity 41.2%; Pred. No. 21;
Matches 7; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Qy 2 NINISLPSVYPDQKSLNLY 18
Db 148 NVNVTPLSHFPDNEINLE 164

RESULT 11
F98050
lysozyme (EC 3.2.1.17) [imported] - *Streptococcus pneumoniae* (strain R6)
C;Species: *Streptococcus pneumoniae*
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C;Accession: F98050
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
V, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: F98050
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-501 <YUR>
A;Cross-references: GB:AE007317; PIDN:AAL00235.1; PID:G15459086; GSPDB:GN00174
C;Genetics:
A;Gene: *lytC*
C;Keywords: glycosidase; hydrolase

Query Match 42.2%; Score 46; DB 2; Length 501;
Best Local Similarity 45.0%; Pred. No. 31;
Matches 9; Conservative 4; Mismatches 3; Indels 4; Gaps 1;

Qy 1 YNINISLPSVYPDQKSLNLY 20
Db 385 YNMNLGYPIYV---DVENVN 400

RESULT 12
A44358
zyxin - chicken
C;Species: *Gallus gallus* (chicken)
C;Date: 10-Jun-1993 #sequence_revision 06-Feb-1995 #text_change 21-Jul-2000
C;Accession: A44358; S30506
R;Sadler, I.; Crawford, A.W.; Michelsen, J.W.; Beckerle, M.C.
J. Cell Biol. 119, 1573-1587, 1992
A;Title: Zyxin and cCRP: two interactive LIM domain proteins associated with the cytoske
A;Reference number: A44358; MUID:93107157; PMID:1469049
A;Accession: A44358
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-542 <SAD>
A;Cross-references: EMBL:X69190; NID:G63897; PIDN:CAA48936.1; PID:G63898
A;Note: sequence extracted from NCBI backbone (NCBI:121172; NCBI:121174)
C;Superfamily: LIM metal-binding repeat homology
F;352-404/Domain: LIM metal-binding repeat homology <LIM1>
F;412-463/Domain: LIM metal-binding repeat homology <LIM2>
F;472-533/Domain: LIM metal-binding repeat homology <LIM3>

Query Match 42.2%; Score 46; DB 2; Length 542;
Best Local Similarity 64.3%; Pred. No. 34;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 NINISLPSYYPDQK 15
:|||||:|:
Db 16 SINISTPSFYNPQK 29

RESULT 13

H81670

conserved hypothetical protein TC0741 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: H81670
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: H81670
A:Status: preliminary
A:Genetics:
A:Molecule type: DNA
A:Residues: 1-1007 <TET>
A:Cross-references: GB:AE002342; GB:AE002160; NID:g7190763; PIDN:AAF39550.1; PID:g719077
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0741

Query Match 42.2%; Score 46; DB 2; Length 1007;
Best Local Similarity 69.2%; Pred. No. 70;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 NISLPSYYPDQKS 16
:|||||:
Db 220 NISLPSYSPTDKS 232

RESULT 14

G81699

peptide ABC transporter, permease protein TC0473 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C:Accession: G81699
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: G81699
A:Status: preliminary
A:Genetics:
A:Molecule type: DNA
A:Residues: 1-287 <TET>
A:Cross-references: GB:AE002315; GB:AE002160; NID:g7190506; PIDN:AAF39321.1; PID:g719051
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0473
C:Superfamily: oligopeptide permease protein oppB

Query Match 41.7%; Score 45.5; DB 2; Length 287;
Best Local Similarity 58.8%; Pred. No. 19;
Matches 10; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 5 ISLPSYYPD-QKSLENY 20
:|||||:|:
Db 40 IFPLSLYPDYKTFPNY 56

RESULT 15

T28119

hypothetical protein ZK945.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T28119
R:Wilkinson, J.
submitted to the EMBL Data Library, March 1995
A:Reference number: Z20472

A:Accession: T28119
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-260 <WIL>
A:Cross-references: EMBL:Z48544; NID:g695493; PIDN:CAA88436.1; GSPDB:GN00020; CBSP:ZK9
A:Experimental source: clone ZK945
C:Genetics:
A:Gene: CBSP:ZK945.2
A:Map position: 2
A:Introns: 24/1; 61/3
C:Superfamily: multicatalytic endopeptidase complex chain C9

Query Match 41.3%; Score 45; DB 2; Length 260;
Best Local Similarity 36.8%; Pred. No. 21;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 2 NINISLPSYYPDQKSLENY 20
:|||||:|:
Db 73 NVGVAVAGNYPDGFALKNY 91

Search completed: March 10, 2004, 12:12:31
Job time : 11.7692 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:51:36 ; Search time 6.15385 Seconds
(without alignments)
169.228 Million cell updates/sec

Title: US-10-044-703-72

Perfect score: 109

Sequence: 1 YNINISLPSPYDPQKSLENY 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	109	100.0	228	1 MP64 MYCTU	P19996 mycobacteri
2	49	45.0	1433	1 SUBF_BACSU	P16397 bacillus su
3	47	43.1	429	1 TRB2_METH	O27520 methanobact
4	47	43.1	432	1 ARQA_STAAM	Q99u25 staphylococ
5	47	43.1	432	1 ARQA_STAAP	Q8csl1 staphylococ
6	47	43.1	433	1 ARQA_STAEP	Q8csl1 staphylococ
7	47	43.1	1272	1 UBP2_YEAST	Q01476 saccharomyc
8	46	42.2	542	1 ZVX_CHICK	Q04584 gallus gall
9	46	42.2	1007	1 Y7A1_CHLMU	Q9pjt6 chlamydia m
10	45.5	41.7	442	1 TOLB_SHEON	Q8edj6 shewanella
11	45	41.3	259	1 PSA3_CABEL	Q99583 caenorhabdi
12	45	41.3	275	1 IF2A_PYRAB	Q9v0e4 pyrococcus
13	45	41.3	275	1 IF2A_PYRPU	Q8u1r5 pyrococcus
14	45	41.3	275	1 IF2A_PYRHO	O58655 pyrococcus
15	45	41.3	520	1 HWC5_CRIGR	P13704 cricetulus
16	45	41.3	520	1 HWC5_HUMAN	Q01581 homo sapien
17	45	41.3	520	1 HWC5_RAT	P17425 rattus norv
18	45	41.3	522	1 HWC5_CHICK	P23228 gallus gall
19	45	41.3	575	1 CNXG_RAT	Q54888 rattus norv
20	45	41.3	1135	1 RPA2_RAT	O54888 rattus norv
21	44.5	40.8	148	1 YXK3_ODOSI	P49829 odontella s
22	44	40.4	420	1 O63A_DROME	Q9vzw8 drosophila
23	44	40.4	430	1 ARQA_STAUA	Q05615 staphylococ
24	44	40.4	455	1 HWC2_BLAG	P54870 blattella g
25	44	40.4	1037	1 ATC3_SCHPO	P22189 schizosacch
26	44	40.4	1135	1 RPA2_MOUSE	P70700 mus musculu
27	43.5	39.9	919	1 RPO2_CAPVK	P16716 capripoxvir
28	43.5	39.9	1161	1 RPO2_FOWEV	Q91544 fowlpox vir
29	43	39.4	153	1 YH17_YEAST	P38898 saccharomyc
30	43	39.4	237	1 Y29A_METYA	P13133 methanococc
31	43	39.4	382	1 ALBU_FANCA	P18447 rana catesb
32	43	39.4	414	1 IDHC_MICME	Q922k9 microtus me
33	43	39.4	414	1 IDHC_MICOH	Q922k8 microtus oc

RESULT 1

MP64 MYCTU
ID_MP64 MYCTU STANDARD; PRT; 228 AA.
AC P19996; 1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Immunogenic protein MPT64/MP64 precursor (Antigen MPT64/MP64).
GN MPT64 OR MP64 OR RV1980C OR MT2032 OR MTCY39.39 OR MB2002C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV / 27294;
RX MEDLINE=94222581; PubMed=7513311;
RA Oettinger T., Andersen A.B.;
RT "Cloning and B-cell-epitope mapping of MPT64 from Mycobacterium tuberculosis H37RV.";
RL Infect. Immun. 62:2058-2064 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=Erdmann;
RX MEDLINE=97313466; PubMed=9169770;
RA Harth G., Lee B.Y., Horwitz M.A.;
RT "High-level heterologous expression and secretion in rapidly growing nonpathogenic mycobacteria of four major Mycobacterium tuberculosis extracellular proteins considered to be leading vaccine candidates and drug targets.";
RT Infect. Immun. 65:2321-2328 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Biglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L., Delcher A., Utterback T., Weidman J., Knouri H., Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;

ALIGNMENTS

34 43 39.4 414 1 IDHC_MOUSE
35 43 39.4 414 1 IDHC_RAT
36 42.5 39.0 395 1 NARK_BACSU
37 42 38.5 181 1 Y095_BORBU
38 42 38.5 266 1 XYN_CALSA
39 42 38.5 202 1 YAO3_SCHPO
40 42 38.5 407 1 WCAI_ECOLI
41 42 38.5 431 1 TOLB_WIGBR
42 42 38.5 709 1 SYT_ARATH
43 42 38.5 1036 1 SYI_CHLTR
44 42 38.5 1079 1 RPA2_HUMAN
45 41.5 38.1 1451 1 DPOA_RAT

O88844 mus musculu
P41562 rattus norv
P46907 bacillus su
O51122 borrelia bu
P23553 caldocellum
Q10082 schizosacch
P32057 escherichia
Q8d2e2 wiggleswort
O4630 arabidopsis
O84022 chlamydia t
Q919y6 homo sapien
O89042 rattus norv

RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RL laboratory strains.";
 RN J. Bacteriol. 184:5479-5490 (2002).
 RN [5]

RP SEQUENCE FROM N.A.
 RC SPECIES=M.Dovis; STRAIN=BCG;
 RX MEDLINE=89079301; PubMed=2642468;
 RA Yamaguchi R., Matsuo K., Yamazaki A., Abe C., Nagai S., Terasaka K.,
 RA Yamada T.;
 RT "Cloning and characterization of the gene for immunogenic protein
 RT MPB64 of Mycobacterium bovis BCG.";
 RL Infect. Immun. 57:283-288 (1989).
 RN [6]

RP SEQUENCE FROM N.A.
 RC SPECIES=M.Dovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Dutfoy S., Grondin J., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
 RN [7]

RP SEQUENCE OF 24-64; 89-119 AND 137-164.
 RC SPECIES=M.tuberculosis;
 RX MEDLINE=92366936; PubMed=1502498;
 RA Wiker H.G., Nagai S., Harboe M., Ljungqvist L.;
 RT "A family of cross-reacting proteins secreted by Mycobacterium
 RL tuberculosis.";
 RL Scand. J. Immunol. 36:307-319 (1992).
 RN [8]

RP SEQUENCE OF 24-53.
 RC SPECIES=M.bovis;
 RX MEDLINE=86166829; PubMed=3514457;
 RA Harboe M., Nagai S., Patarroyo M.E., Torres M., Ramirez C., Cruz N.;
 RT "Properties of proteins MPB64, MPB70, and MPB80 of Mycobacterium
 RL bovis BCG.";
 RL Infect. Immun. 52:293-302 (1986).
 RN [9]

CC -!- SUBCELLULAR LOCATION: Secreted.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X75361; CAAS3143.1; -;
 DR EMBL; U82235; AAB61538.1; -;
 DR EMBL; Z74025; CA98382.1; -;
 DR EMBL; A3007056; AAK46307.1; ALT_INIT.
 DR EMBL; U34849; AAC44034.1; -;
 DR EMBL; BX248340; CAD94704.1; -;
 DR PIR; A30545; A30545.
 DR PIR; B70756; B70756.
 DR TIGR; MT2032; -;
 DR Truncuist; Rv1980c; -;
 KW Antigen; Signal; Complete proteome.
 FT SIGNAL 1 23
 FT CHAIN 24 228 IMMUNOGENIC PROTEIN MPT64/MPB64.
 FT CONFLICT 29 29 C -> H OR R (IN REF. 8).
 FT CONFLICT 41 41 C -> Y (IN REF. 8).
 FT CONFLICT 41 41 C -> Y (IN REF. 8).
 FT SEQUENCE 228 AA; 24855 MW; D79850CD1907999C CRC64;
 Query Match 100.0%; Score 109; DB 1; Length 228;
 Best Local Similarity 100.0%; Pred. No. 6.6e-10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YNINISLPSPYDPQKSLNLY 20
 DB 50 YNINISLPSPYDPQKSLNLY 69

RESULT 2
 SUBF_BACSU STANDARD; PRT; 1433 AA.
 AC P16357;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Bacillopeptidase F precursor (EC 3.4.21.-) (Esterase) (RP-I protease)
 DE (90 kDa serine proteinase).
 GN BPR OR BPR OR BSU15300.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OC NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 195-222.
 RX MEDLINE=90170864; PubMed=2106512;
 RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
 RA Pero J.;
 RT "Bacillopeptidase F of Bacillus subtilis: purification of the protein
 RT and cloning of the gene.";
 RL J. Bacteriol. 172:1470-1477 (1990).
 RN [2]
 RN REVISIONS.
 RC STRAIN=168;
 RX MEDLINE=90368623; PubMed=2118514;
 RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
 RA Pero J.;
 RL J. Bacteriol. 172:5520-5521 (1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90216713; PubMed=2108961;
 Wu X.-C., Nathoo S., Pang A.S.-H., Carne T., Wang S.-L.;
 RT "Cloning, genetic organization, and characterization of a structural
 RT gene encoding bacillopeptidase F from Bacillus subtilis.";
 RL J. Biol. Chem. 265:6845-6850 (1990).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriello S., Bruschini L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschini L., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conner I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Haeleth J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holstappel S., Hooson S., Hulio M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaere-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Roche E., Roche M., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Toato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256 (1997).
 RN [5]
 RP SEQUENCE OF 1-211 FROM N.A.

```

RX MEDLINE=89008108; PubMed=3139638;
RA Beall B., Lowe M., Lutkenhaus J.;
RT "Cloning and Characterization of Bacillus subtilis homologs of
RL Escherichia coli cell division genes ftsZ and ftsA.";
RN J. Bacteriol. 170:4855-4864(1998).
[6]
RN SEQUENCE OF 1410-1433 FROM N.A.
RP STRAIN=168 / Marburg;
RC MEDLINE=90174995; PubMed=2106671;
RX Masuda E.S., Anaguchi H., Sato T., Takeuchi M., Kobayashi Y.;
RA "Nucleotide sequence of the sporulation gene spoIIGA from Bacillus
RT subtilis.";
RL Nucleic Acids Res. 18:657-657(1990).
[7]
RN SEQUENCE OF 195-219.
RP STRAIN=Natto 16;
RC Kato T., Yamagata Y., Arai T., Ichishima E.;
RT "Purification of a new extracellular 90-kDa serine proteinase with
RL isoelectric point of 3.9 from Bacillus subtilis (natto) and
RL elucidation of its distinct mode of action.";
RN Biosci. Biotechnol. Biochem. 56:1166-1169(1992).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M29035; AAA62679.1; -
CC EMBL; J05400; AAA83362.1; -
CC EMBL; Z91111; CAB13403.1; -
CC EMBL; M22630; AAA22458.1; -
CC EMBL; X17344; CAA35224.1; -
CC PIR; A36734; A36734.
CC HSP; P00782; 28BT.
CC MEROPS; S08.017; -.
CC Subtilisin; BG10233; bpr.
CC InterPro; IPR008757; Peptidase M6.
CC InterPro; IPR000209; Peptidase S8.
CC InterPro; IPR009020; Protease inhib.
CC Pfam; PF05547; Peptidase M6; 1.
CC Pfam; PF00082; Peptidase S8; 1.
CC PRINTS; P00723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE ASP; FALSE NEG.
CC PROSITE; PS00137; SUBTILASE HIS; 1.
CC PROSITE; PS00138; SUBTILASE SER; 1.
CC KW Hydrolyase; Serine protease; Zymogen; Signal; Complete proteome.
CC
CC -----
CC SIGNAL 1 30
CC PROPEP 31 194 POTENTIAL.
CC CHAIN 195 755 BACILLOPEPTIDASE F.
CC PROPEP 756 1433 POTENTIAL.
CC FT ACT SITE 227 227 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT SITE 274 274 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT SITE 452 452 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT CONFLICT 219 219 T -> A (IN REF. 7).
CC FT CONFLICT 393 393 A -> V (IN REF. 3).
CC FT CONFLICT 829 834 KQNKKA -> N (IN REF. 3).
CC FT CONFLICT 836 841 QGVQLP -> RTRLKS (IN REF. 3).
CC FT CONFLICT 844 852 AQSVVETG -> FCRSRHVS (IN REF. 3).
CC FT CONFLICT 853 1433 MISSING (IN REF. 3).
CC SQ SEQUENCE 1433 AA; 154577 MW; 98DF6846897807C9 CRC64;
Query Match 45.0%; Score 49; DB 1; Length 1433;
Best Local Similarity 44.4%; Pred. No. 15;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
QY 1 YNINISLPSPYDPQKSLK 18
DB 957 YTIKVAAPGYSDSEFSVE 974

```

```

RESULT 3
TRB2_METTH
ID TRB2_METTH STANDARD; PRT; 429 AA.
AC O27520;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tryptophan synthase beta chain 2 (EC 4.2.1.20).
GN TRPB2 OR MTH1476.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-L., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RL J. Bacteriol. 179:7135-7155(1997).
CC -1- FUNCTION: The beta subunit is responsible for the synthesis of L-
CC tryptophan from indole and L-serine.
CC -1- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
CC = L-tryptophan + glyceraldehyde 3-phosphate.
CC -1- COFACTOR: Pyridoxal phosphate (By similarity).
CC -1- PATHWAY: Tryptophan biosynthesis; fifth (last) step.
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC similarity).
CC -1- SIMILARITY: Belongs to the trpB family.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AB000908; AAB85951.1; -
CC PIR; F69063; F69063.
CC HSP; P00933; 2WSY.
CC HAMAP; MF 00133; -; 1.
CC InterPro; IPR001926; B6 enzyme beta.
CC InterPro; IPR006316; Trp synth_b like.
CC InterPro; IPR006653; Trp synth_b_rel.
CC Pfam; PF00291; PALP; 1.
CC TIGRFAMs; TIGR01415; trpB_rel; 1.
CC PROSITE; PS00169; TRP_SYNTHASE_BETA; 1.
CC KW Tryptophan biosynthesis; Pyridoxal phosphate; Lyase;
CC Complete proteome.
CC FT BINDING 110 110 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC SQ SEQUENCE 429 AA; 47487 MW; 0359847317C047DD CRC64;
Query Match 43.1%; Score 47; DB 1; Length 429;
Best Local Similarity 55.6%; Pred. No. 7.8;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 1 YNINISLPSPYDPQKSLK 18
DB 17 YNINFDLPSPLPBPKNPE 34

```

```

RESULT 4
AROA_STAA

```

```

ID AROA STAAW STANDARD; PRT; 432 AA.
AC Q99U25;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
DE enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN AROA OR SAVI464 OR SA1297.
OS Staphylococcus aureus (strain N315).
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418446;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizukami U., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the EPSP synthase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP003362; BAB57626.1; -
CC EMBL; AP003362; BAB57626.1; -
CC PIR; H89924; H89924.
CC HAMAP; MF 00210; -; 1.
CC InterPro; IPR006264; AROA.
CC Pfam; PF00275; EPSP_synthase; 1.
CC ProDom; PD001867; EPSP_synthase; 1.
CC TIGRFAMS; TIGR01356; aroA; 1.
CC PROSITE; PS00104; EPSP_SYNTHASE_1; 1.
CC PROSITE; PS00885; EPSP_SYNTHASE_2; 1.
CC Aromatic amino acid biosynthesis; transferase; Complete proteome.
CC KW SEQUENCE 432 AA; 47068 MW; A42102057AD15C72 CRC64;
CC -----
Query Match 43.1%; Score 47; DB 1; Length 432;
Best Local Similarity 47.1%; Pred. No. 7.9;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 3 INISLPSYDPQKSLEN 19
Db 414 VNVSPFGFLPKLLEN 430

RESULT 5
ID AROA STAAW STANDARD; PRT; 432 AA.
AC GNWNS;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
DE enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN AROA OR SE1153.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";

```

```

DE enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN AROA OR MW1354.
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=196620;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwano N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827(2002).
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the EPSP synthase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP004827; BAB95219.1; -
CC HAMAP; MF 00210; -; 1.
CC InterPro; IPR006264; AROA.
CC Pfam; PF00275; EPSP_synthase; 1.
CC ProDom; PD001867; EPSP_synthase; 1.
CC TIGRFAMS; TIGR01356; aroA; 1.
CC PROSITE; PS00104; EPSP_SYNTHASE_1; 1.
CC PROSITE; PS00885; EPSP_SYNTHASE_2; 1.
CC Aromatic amino acid biosynthesis; transferase; Complete proteome.
CC KW SEQUENCE 432 AA; 47003 MW; 7C092783DC8E3F5F CRC64;
CC -----
Query Match 43.1%; Score 47; DB 1; Length 432;
Best Local Similarity 47.1%; Pred. No. 7.9;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 3 INISLPSYDPQKSLEN 19
Db 414 VNVSPFGFLPKLLEN 430

RESULT 6
ID AROA STAAW STANDARD; PRT; 433 AA.
AC Q8CS11;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
DE enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN AROA OR SE1153.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RX PubMed=12950922;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";

```

RL Mol. Microbiol. 49:1577-1593 (2003).
 CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
 CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
 CC -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
 CC sixth step.
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -!- SIMILARITY: Belongs to the EPP synthase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE016747; AA004750.1; -;
 CC HAVAP; MF_00210; -; 1.
 CC InterPro; IPR006264; AroA.
 CC InterPro; IPR001986; EPP synth.
 CC Pfam; PF00275; EPP synthase; 1.
 CC ProDom; PD001867; EPP synthase; 1.
 CC TIGRFAMs; TIGR01356; aroA; 1.
 CC PROSITE; PS00104; EPP SYNTHASE 1; 1.
 CC PROSITE; PS00885; EPP SYNTHASE_2; 1.
 CC Aromatic amino acid biosynthesis; Transferase; Complete proteome.
 KW Aromatic amino acid biosynthesis; Transferase; Complete proteome.
 SQ SEQUENCE 433 AA; 47003 MW; 880036E881BB76AD CRC64;
 Query Match 43.1%; Score 47; DB 1; Length 433;
 Best Local Similarity 47.1%; Pred. No. 7.9;
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 3 INISLPSYYPDQKSLN 19
 Db 414 VNVSPFGFLPKLLEN 430
 RESULT 7
 ID UBP2_YEAST STANDARD; PRT; 1272 AA.
 AC Q01476; Q99357;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 2 (EC 3.1.2.15) (Ubiquitin
 DE thiolesterase 2) (Ubiquitin-specific processing protease 2)
 DE (Deubiquitinating enzyme-2).
 GN UBP2 OR YOR124C OR O3281 OR YOR3281C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054674; PubMed=1429680;
 RA Baker R.T., Tobias J.W., Varshavsky A.;
 RT "Ubiquitin-specific proteases of Saccharomyces cerevisiae. Cloning of
 RT UBP2 and UBP3, and functional analysis of the UBP gene family.";
 RL J. Biol. Chem. 267:23364-23375 (1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / FY1679;
 RX MEDLINE=97060020; PubMed=8904341;
 RA Wiemann S., Rechmann S., Benes V., Voss H., Schwager C., Vleck C.,
 RA Stegmann J., Zimmermann J., Exfle H., Paces V., Ansorge W.;
 RT "Sequencing and analysis of 51 kb on the right arm of chromosome XV
 RT from Saccharomyces cerevisiae reveals 30 open reading frames.";
 RL Yeast 12:281-288 (1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97344368; PubMed=9200815;
 RA Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C.,

RA Schwager C., Paces V., Sander C., Ansorge W.;
 RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV.";
 RL Yeast 13:655-672 (1997).
 CC -!- FUNCTION: Has an ATP-independent isopeptidase activity, cleaving
 CC at the carboxyl terminus of the ubiquitin moiety in natural or
 CC engineered linear fusion proteins, irrespective of their size or
 CC the presence of an amino-terminal extension to ubiquitin.
 CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -!- SIMILARITY: Belongs to peptidase family C19.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M94916; AA335190.1; -;
 CC EMBL; X94335; CAA64043.1; -;
 CC EMBL; X90518; CAA62120.1; -;
 CC EMBL; Z75032; CAA9323.1; -;
 CC PIR; S60999; S60999.
 CC GerMOnline; 143712; -;
 CC MEROPS; C19.003; -;
 CC SGD; S0005650; UBP2.
 CC GO; GO:0005737; Cytoplasm; IC.
 CC GO; GO:0004843; Fubiquitin-specific protease activity; IDA.
 CC InterPro; IPR001394; Peptidase_C19.
 CC Pfam; PF00443; UCH; 1.
 CC PROSITE; PS00972; UCH_2_1; 1.
 CC PROSITE; PS00973; UCH_2_2; 1.
 CC PROSITE; PS0235; UCH_2_3; 1.
 KW Ubi conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 FT ACT_SITE 745 745 BY SIMILARITY.
 FT ACT_SITE 1201 1201 BY SIMILARITY.
 FT ACT_SITE 1209 1209 BY SIMILARITY.
 FT CONFLICT 658 665 MISSING (IN REF. 1).
 SQ SEQUENCE 1272 AA; 146354 MW; 6D106539AE5C5F3F CRC64;
 Query Match 43.1%; Score 47; DB 1; Length 1272;
 Best Local Similarity 44.4%; Pred. No. 27;
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
 QY 3 INISLPSYYPDQKSLN 20
 Db 449 INLSVSHYVTRDIRNY 466
 RESULT 8
 ID ZYX_CHICK STANDARD; PRT; 542 AA.
 AC Q04584;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Zyxin.
 DE Zyxin.
 GN Zyx.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93107157; PubMed=1469049;
 RA Sadler I., Crawford A.W., Michelsen J.W., Beckerle M.C.;
 RT "Zyxin and cCRP: two interactive LIM domain proteins associated with
 RT the cytoskeleton.";
 RL J. Cell Biol. 119:1573-1587 (1992).
 CC -!- FUNCTION: Adhesion plaque protein. Binds alpha-actinin and the CRP
 CC protein. May be a component of a signal transduction pathway that

```
CC mediates adhesion-stimulated changes in gene expression.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; associates with the actin
CC cytoskeleton near the adhesion plaques.
CC -1- SIMILARITY: Contains 3 LIM zinc-binding domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; AE002342; AAP39550.1; -
CC PIR; H81670; H81670.
CC TIGR; TC0741; -
CC KW Hypothetical protein; Signal; Complete proteome.
CC FT SIGNAL 1 51 POTENTIAL.
CC FT CHAIN 52 1007 HYPOTHETICAL PROTEIN TC0741.
CC SQ SEQUENCE 1007 AA; 104006 MW; 842800C0871B1518 CRC64;
CC
CC Query Match 42.2%; Score 46; DB 1; Length 1007;
CC Best Local Similarity 69.2%; Pred. No. 30;
CC Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CC
CC QY 4 NISLPSYPDPQKS 16
CC Db 220 NMSLPSYSPTDKS 232
CC
CC RESULT 10
CC TOLB_SHEON STANDARD; PRT; 442 AA.
CC ID TOLB_SHEON Q8EDJ9;
CC AC Q8EDJ9;
CC DT 10-OCT-2003 (Rel. 42, Created)
CC DT 10-OCT-2003 (Rel. 42, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE TOLB protein precursor.
CC GN TOLB OR SO2748.
CC OS Shewanella oneidensis.
CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
CC Alteromonadaceae; Shewanella.
CC NCBI_TaxID=70863;
CC [1]
CC RN SEQUENCE FROM N.A.
CC RP STRAIN=MR-1.
CC RC MEDLINE=2297686; PubMed=12368813;
CC RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
CC Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
CC Meyer T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,
CC DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
CC Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
CC Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
CC Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
CC Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
CC "Genome sequence of the dissimilatory metal ion-reducing bacterium
CC Shewanella oneidensis";
CC Nat. Biotechnol. 20:1118-1123(2002).
CC -1- FUNCTION: Involved in the tonB-independent uptake of proteins (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: Belongs to the tolB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

QY 1 YNINISLPSPYDPQKSL 17
DB 299 YSIDTE-PSWYPDGKSL 314

RESULT 11
PSA3_CABEL STANDARD; PRT; 259 AA.
AC Q09583;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Proteasome subunit alpha type 3 (EC 3.4.25.1) (Proteasome subunit
alpha 7).
GN PAS-7 OR ZK945.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilkinson-Sproat J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: The proteasome is a multicatalytic proteinase complex
CC which is characterized by its ability to cleave peptides with Arg,
CC phe, Tyr, Leu, and Glu adjacent to the leaving group at neutral or
CC slightly basic pH. The proteasome has an ATP-dependent proteolytic
CC activity (By similarity).
CC -!- CATALYTIC ACTIVITY: Cleavage at peptide bonds with very broad
CC specificity.
CC -!- PATHWAY: Involved in an ATP/ubiquitin-dependent non-lysosomal
CC proteolytic pathway.
CC -!- SUBUNIT: The proteasome is composed of at least 15 non identical
CC subunits which form a highly ordered ring-shaped structure (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family T1A.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z48544; CAB98436.1; -
CC FIR; T28119; T28119.
CC HSP; P21242; ILYP.
CC MEROPS; T01.977; -.
CC WormPep; ZK945.2; CE01733.
CC InterPro; IPR000426; Pept_T1 subA.
CC InterPro; IPR001353; Peptidase_T1.
CC Pfam; PF00227; Proteasome; 1.
CC PROSITE; PS00388; PROTEASOME_A; 1.
CC Proteasome; Hydrolase; Protease; Threonine protease.
CC INIT MET 0 BY SIMILARITY.
CC SEQUENCE 259 AA; 28805 MW; B2B52A353E293452 CRC64;

Query Match 41.3%; Score 45; DB 1; Length 259;
Best Local Similarity 36.8%; Pred. No. 9.1;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 2 NINISLPSPYDPQKSL 20
DB 72 NVGVAVAGNYPDGFKNY 90

RESULT 12
IF2A_PVRAB STANDARD; PRT; 275 AA.
AC Q9V0E4;

QY 1 YNINISLPSPYDPQKSL 19
DB 224 YRIDTAPDYKAEVL 242

RESULT 13
IF2A_PVRFU STANDARD; PRT; 275 AA.
AC Q8UIE5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Probable translation initiation factor 2 alpha subunit (eIF-2-alpha).
GN EIF2A OR PF1140.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.

Query Match 41.3%; Score 45; DB 1; Length 275;
Best Local Similarity 42.1%; Pred. No. 9.7;
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 YNINISLPSPYDPQKSL 19
DB 224 YRIDTAPDYKAEVL 242

RESULT 13
IF2A_PVRFU STANDARD; PRT; 275 AA.
AC Q8UIE5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Probable translation initiation factor 2 alpha subunit (eIF-2-alpha).
GN EIF2A OR PF1140.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.

```


[illegible]

DR EMBL; L00334; AAA37076.1; --
DR EMBL; L00326; AAA37076.1; JOINED.
DR EMBL; L00327; AAA37076.1; JOINED.
DR EMBL; L00328; AAA37076.1; JOINED.
DR EMBL; L00329; AAA37076.1; JOINED.
DR EMBL; L00330; AAA37076.1; JOINED.
DR EMBL; L00331; AAA37076.1; JOINED.
DR EMBL; L00332; AAA37076.1; JOINED.
DR EMBL; L00333; AAA37076.1; JOINED.
DR PIR; A25332; A25332.
DR InterPro; IPR000590; HMG_CoA_synth_AS.
DR InterPro; IPR008260; HMG_CoA_synth.
DR Pfam; PF01154; HMG_CoA_synth; 1.
DR PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.
KW Transferase; Cholesterol biosynthesis; Multigene family.
FT ACT_SITE 129 129 POTENTIAL.
SQ SEQUENCE 520 AA; 57318 MW; 2D4CAAETDEE5D6BB CRC64;

Query Match 41.3%; Score 45; DB 1; Length 520;
Best Local Similarity 50.0%; Pred.No. 20;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Qy 3 INTSLPSYYPDQKSLNLY 20
Db 22 LEIYFSPQYVDQAELEKY 39

Search completed: March 10, 2004, 12:06:34
Job time : 7.15385 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:57:36 ; Search time 31.6667 Seconds
(without alignments)
199.275 Million cell updates/sec

Title: US-10-044-703-72

Perfect score: 109
Sequence: 1 YNINISLPSYPDQKSLNLY 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phase: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_rvirus: *
16: sp_bacteriaph: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	95.4	228	2 Q84AL7	Q84AL7 mycobacteri
2	51	46.8	227	16 Q7TXE4	Q7TXE4 mycobacteri
3	51	46.8	938	5 Q8IEB6	Q8IEB6 plasmodium
4	51	46.8	3401	5 Q8IB03	Q8IB03 plasmodium
5	50	45.9	227	16 Q53283	Q53283 mycobacteri
6	49	45.0	117	8 Q8SKK8	Q8SKK8 brachistost
7	49	45.0	117	8 Q85CT1	Q85CT1 brachistost
8	49	45.0	271	16 Q7VRI9	Q7VRI9 candidatus
9	49	45.0	331	5 Q965L6	Q965L6 caenorhabdi
10	49	45.0	442	5 Q8MSX1	Q8MSX1 drosophila
11	49	45.0	525	5 Q8VJ47	Q8VJ47 drosophila
12	49	45.0	555	5 Q8II09	Q8II09 plasmodium
13	49	45.0	906	5 Q8IDD0	Q8IDD0 plasmodium
14	49	45.0	1433	2 Q456L6	Q456L6 bacillus su
15	49	45.0	1548	5 Q9NGQ6	Q9NGQ6 drosophila
16	49	45.0	1818	5 Q9VZM1	Q9VZM1 drosophila

17	48.5	44.5	631	2 P71143	P71143 clostridium
18	48	44.0	553	5 Q811H7	Q811H7 drosophila
19	48	44.0	555	5 Q9VVB4	Q9VVB4 drosophila
20	48	44.0	1005	5 Q8IDB9	Q8IDB9 plasmodium
21	48	44.0	1124	17 Q9HKQ3	Q9HKQ3 thermoplas
22	47	43.1	103	2 Q8ZEH7	Q8ZEH7 enterococcu
23	47	43.1	103	2 Q8ZEH7	Q8ZEH7 enterococcu
24	47	43.1	103	2 Q8ZEH7	Q8ZEH7 enterococcu
25	46.5	42.7	1342	5 Q9N3Y8	Q9N3Y8 caenorhabdi
26	46.5	42.7	1342	5 Q9N3Y8	Q9N3Y8 caenorhabdi
27	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
28	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
29	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
30	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
31	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
32	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
33	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
34	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
35	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
36	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
37	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
38	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
39	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
40	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
41	46	42.2	1401	5 Q8MMN1	Q8MMN1 dictyosteli
42	45.5	41.7	279	10 Q7X9H7	Q7X9H7 arabidopsis
43	45.5	41.7	287	16 Q9PKJ2	Q9PKJ2 chlamydia m
44	45.5	41.7	289	10 Q9FKK2	Q9FKK2 arabidopsis
45	45.5	41.7	324	16 Q8DIU7	Q8DIU7 wiggleswort

ALIGNMENTS

RESULT 1

Q84AL7 PRELIMINARY; PRT; 228 AA.
ID Q84AL7
AC Q84AL7;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Secreted protein Mpt64.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee C.F.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY208674; AAC48435.1; -
SQ SEQUENCE 228 AA; 24868 MW; 7E1106C2B39F41F7 CRC64;

Query Match 95.4%; Score 104; DB 2; Length 228;
Best Local Similarity 95.0%; Pred. No. 4.9e-06;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YNINISLPSYPDQKSLNLY 20

DB 50 YNINISLPSYPDQKSLNLY 69

RESULT 2

Q7TXE4 PRELIMINARY; PRT; 227 AA.
ID Q7TXE4
AC Q7TXE4;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Probable conserved secreted protein TB22.2.
GN TB22.2 OR MB3062C.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Query Match 45.0%; Score 49; DB 8; Length 117;
Best Local Similarity 44.4%; Pred.No. 8.2;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 INISLPSYYDPQKSLNLY 20
DB 21 VGLGLPSIYPDNEKL SAY 38

RESULT 8
Q7VRI9 PRELIMINARY; PRT; 271 AA.
AC Q7VRI9;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Uridine phosphorylase (EC 2.4.2.3).
GN UDP OR BFL624.
OS Candidatus Blochmannia floridanus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; ant endosymbionts; Candidatus Blochmannia.
OX NCBI_TaxID=203907;
RN [1]
SEQUENCE FROM N.A.

RX MEDLINE=22784745; PubMed=12886019;
RA Gil R., Silva F.J., Zientz E., Delmotte F., Gonzalez-Candelas F.,
RA Latorre A., Rausell C., Kamerbeek J., Gadau J., Hoelldobler B.,
van Ham R.C.H.J., Gross R., Moya A.;
RT "the genome sequence of Blochmannia floridanus: comparative analysis
of reduced genomes";
RL Proc. Natl. Acad. Sci. U.S.A. 100:9388-9393(2003).
DR EMBL; BX248586; CAD83299.1;-;
KW Glycosyltransferase; Transferrase; Complete proteome.
SQ SEQUENCE 271 AA; 29799 MW; 17F2E4A7EAF5C598 CRC64;

Query Match 45.0%; Score 49; DB 16; Length 271;
Best Local Similarity 40.0%; Pred.No. 20;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 YNINISLPSYYDPQKSLNLY 20
DB 155 WGINVSDFYPGQLNTY 174

RESULT 9
Q96SL6 PRELIMINARY; PRT; 331 AA.
AC Q96SL6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein F34D6.6.
GN F34D6.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RC STRAIN=Bristol N2;
RA Murray J., Wohlmann P., O'Neal D.;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RC STRAIN=Bristol N2;
RA Submitted (SEP-1997) to the EMBL/GenBank/DBDJ databases.
RL [3]
RP SEQUENCE FROM N.A.

```

RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF025454; AAK69371.1; -.
DR WormPep; F34D6.6; CE29870.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004930; F:G-protein coupled receptor activity; IEA.
DR InterPro; IPR003003; 7TM_chemrecept.
DR InterPro; IPR000168; Nm7TM_chemrecept.
DR Pfam; PF01604; 7tm_5; 1.
SQ SEQUENCE 331 AA; 39051 MW; 08ADC8C644523BF5 CRC64;

Query Match 45.0%; Score 49; DB 5; Length 331;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 INISLPSPYDPQKSLNLY 20
Db 165 IRVNLPEYLPQKSLNLF 182

RESULT 10
Q8MSX1 PRELIMINARY; PRT; 442 AA.
AC Q8MSX1;
CT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE L166389.
GN ENC OR CG10847.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophiliidae; Drosophila.
CX NCBI_TaxID=7227;

SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Adams M.D., Celinker S.E., Li P.W., Hoskins R.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Rogers Y.-H.C., Blaziej M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaziej M.D., Zhang Q., Chen L.X.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Besson K.V., Benos P.V., Berwan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RN EMBL; AV118518; AAM49887.1; -.
DR FlyBase; FBgn0004875; enc.
DR GO; GO:0005737; C:cycloplasm; IDA.
DR GO; GO:0007282; P:cystoblast cell division; IMP.
DR GO; GO:0007293; P:egg chamber formation (sensu Insecta); IMP.
DR GO; GO:0007292; P:female gamete generation; IMP.
DR GO; GO:0007294; P:oocyte cell fate determination (sensu Insecta); IMP.
SQ SEQUENCE 442 AA; 43901 MW; 341165A46A47E877 CRC64;

Query Match 45.0%; Score 49; DB 5; Length 442;
Best Local Similarity 50.0%; Pred. No. 33;
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 1 YNINISLPSPYDPQKSLNLY 20
Db 48 YHNSNSTPHYQGNSEGY 67

RESULT 11
Q9VJ47 PRELIMINARY; PRT; 525 AA.
AC Q9VJ47;
CT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE CG17324 protein (GH06505p).
GN BEST:GH06505 OR CG17324.

```

Query Match 45.0%; Score 49; DB 5; Length 525;

Best Local Similarity 52.9%; Pred. No. 39;

Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 4 NISLPSPYDPQKSLNLY 20

|||: |||: |||: |||:

Db 51 NISMYSYFFLEXPVANY 67

RESULT 12

Q8IIC9 PRELIMINARY; PRT; 555 AA.
 ID Q8IIC9; MEDLINE=22255705; PubMed=12368864;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Translation elongation factor EF-1, subunit alpha, putative.
 GN PF11_0245.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=22255705; PubMed=12368864;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
 RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrall B.;
 RT "Genome sequence of the human malaria parasite Plasmodium
 falciparum";
 RL Nature 419:498-511(2002).
 DR EMBL; AB014839; AAN35829.1; ..
 DR GO; GO:0005255; F:GTP binding; IEA.
 DR GO; GO:0003746; F:translation elongation factor activity; IEA.
 DR GO; GO:0008414; P:translational elongation; IEA.
 DR InterPro; IPR004160; EFTU_Cterm.
 DR InterPro; IPR004161; EFTU_D2.
 DR InterPro; IPR000795; EF_GTPbind.
 DR InterPro; IPR009001; Elong init C.
 DR InterPro; IPR009000; Translat_factor.
 DR Pfam; PF00009; GTP_EFTU.1.
 DR Pfam; PF03144; GTP_EFTU_D2; 1.
 DR Pfam; PF03143; GTP_EFTU_D3; 1.
 DR PRINTS; PR00315; ELONGATNFCT.
 KW Elongation factor.
 SQ SEQUENCE 555 AA; 62881 MW; D7B9D98FA90D1625 CRC64;

Query Match 45.0%; Score 49; DB 5; Length 555;
 Best Local Similarity 46.7%; Pred. No. 42;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YNINISLPSYYPDQK 15
 Db 5 PSFNVNAPSYPGMK 19

RESULT 13

Q8IDDO PRELIMINARY; PRT; 906 AA.
 ID Q8IDDO; MEDLINE=22255705; PubMed=12368864;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE DNA helicase.
 GN PF13_0308.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Harris B., Lennart N., Clark L., Line A., Barron A., Corton C.,
 RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
 RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrall B.;

RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL844509; CAD52695.1; ..
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR000330; SNF2 N.
 DR Pfam; PF00271; Helicase C; 1.
 DR Pfam; PF00176; SNF2 N; 1.
 DR SMART; SMO0487; DEXDc; 1.
 DR SMART; SMO0490; HELICc; 1.
 KW Helicase.
 SQ SEQUENCE 906 AA; 106289 MW; 822EBB85929A0FC5 CRC64;

Query Match 45.0%; Score 49; DB 5; Length 906;
 Best Local Similarity 41.3%; Pred. No. 69;
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 4 NISLPSYYPDQKSLNY 20
 Db 124 NVKSPHYNNKINNY 140

RESULT 14

Q45616 PRELIMINARY; PRT; 1433 AA.
 ID Q45616; MEDLINE=96084385; PubMed=8528006;
 AC Q45616; Yanagata Y., Abe R., Fujita Y., Ichishima E.;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE 90k-protease (Bacillopeptidase F) precursor.
 GN HSPN
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=natto;
 RX MEDLINE=96084385; PubMed=8528006;
 RA Yanagata Y., Abe R., Fujita Y., Ichishima E.;
 RT "Molecular cloning and nucleotide sequence of the 90k-serine protease
 Curr. Microbiol. 31:340-344(1995).
 DR EMBL; D44498; BAA07941.1; ..
 DR HSPN; P00782; 2SBT.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR008757; Peptidase M6.
 DR InterPro; IPR000209; Peptidase S8.
 DR Pfam; PF05547; Peptidase M6; 1.
 DR Pfam; PF00082; Peptidase S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 KW Protease; Signal.
 FT SIGNAL 1 30 POTENTIAL.
 SQ SEQUENCE 1433 AA; 154575 MW; 728EF61AF34B0841 CRC64;

Query Match 45.0%; Score 49; DB 2; Length 1433;
 Best Local Similarity 44.4%; Pred. No. 11le+02;
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 1 YNINISLPSYYPDQKSL 18
 Db 957 YTIKVAAPGYSDPSVE 974

RESULT 15

Q9NGQ6 PRELIMINARY; PRT; 1548 AA.
 ID Q9NGQ6
 AC Q9NGQ6;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Cytoplasmic protein encode.
 GN ENC OR CGI0847.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Van Buskirk C., Hawkins N.C., Schubach T.;
 RT "Encore is a member of a novel class of proteins and affects multiple
 RT processes in Drosophila oogenesis.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF243382; AAF68440.1; -
 DR FlyBase: FBgn004875; enc.
 DR GO: GO:0005737; Cytoplasm; IDA.
 DR GO: GO:0007282; P:cystoblast cell division; IMP.
 DR GO: GO:0007293; P:egg chamber formation (sensu Insecta); IMP.
 DR GO: GO:0007292; P:female gamete generation; IMP.
 DR GO: GO:0007294; P:oocyte cell fate determination (sensu Insecta); IMP.
 DR InterPro: IPR001374; R3H.
 DR Pfam: PF01424; R3H; 1.
 SQ SEQUENCE 1548 AA; 164364 MW; F2423CE7E7D5D0CF CRC64;

Query Match 45.0%; Score 49; DB 5; Length 1548;
 Best Local Similarity 50.0%; Pred. NO. 1.2e+02;
 Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 1 YNINISLPSYYPDKSLENY 20
 Db 1280 YHSNSTPHYQQQNSNEGY 1299

Search completed: March 10, 2004, 12:10:59
 Job time : 33.7917 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:50:51 ; Search time 47.5641 Seconds
(without alignments)
118.807 Million cell updates/sec

Title: US-10-044-703-78

Perfect score: 96
Sequence: 1 IDELXTNSLLTSILTYHV 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96	100.0	20	AAE12293	Aae12293 Mycobacte
2	96	100.0	163	AAP91363	Aap91363 MPB-70 pr
3	96	100.0	163	AAW12045	Aaw12045 MPB 70. 4
4	96	100.0	192	AAR07235	Aar07235 Protein an
5	96	100.0	193	AAW99351	Aaw99351 M.tubercu
6	96	100.0	201	AAR06839	Aar06839 Human IL-
7	92	95.8	163	AAR07053	Aar07053 Immunopro
8	76	79.2	220	AAW17979	Aaw17979 Mycobacte
9	76	79.2	220	AAW99350	Aaw99350 M.tubercu
10	71	74.0	231	AAW14905	Aaw14905 Amino aci
11	71	74.0	231	ABW73511	Abw73511 M vaccae
12	71	74.0	231	ABP70878	Abp70878 Mycobacte
13	70	72.9	228	AAW14904	Aaw14904 Amino aci
14	70	72.9	228	ABW73510	Abw73510 M vaccae
15	70	72.9	228	ABP70879	Abp70879 Mycobacte
16	49	51.0	20	AAE12294	Aae12294 Mycobacte
17	47.5	49.5	599	ABG07745	Abg07745 Novel hum
18	47.5	49.5	1605	ABG07744	Abg07744 Novel hum
19	47	49.0	20	AAE12292	Aae12292 Mycobacte
20	46	47.9	180	ABP77905	Abp77905 High leve
21	46	47.9	209	ABP40281	Abp40281 Staphyloc
22	46	47.9	263	ABU43261	Abu43261 Protein e
23	45	46.9	261	AAU33880	Aau33880 Staphyloc
24	45	46.9	263	ABU16138	Abu16138 Protein e
25	45	46.9	266	AAU36698	Aau36698 Staphyloc

ALIGNMENTS

RESULT 1

AAE12293
ID AAE12293 standard; peptide; 20 AA.
XX AC AAE12293;
XX AC
DT 18-DEC-2001 (first entry)
DE Mycobacterium tuberculosis (Mtb) peptide #78.
XX DE
KW Mycobacterium tuberculosis; Mtb peptide; antibacterial; vaccine;
infection; anti-Mtb immune response.
XX KW
OS Mycobacterium tuberculosis.
XX OS
PW WO200170774-A2.
XX PW
PD 27-SEP-2001.
XX PD
PF 20-MAR-2001; 2001WO-US008906.
XX PF
PR 20-MAR-2000; 2000US-0190834P.
XX PR
PA (UYBR-) UNIV BROWN RES FOUND.
XX PA
PI Degroot AS;
XX PI
DR WPI; 2001-616401/71.
XX DR
PT New vaccine for immunizing a mammalian subject, preferably humans,
against infection caused by Mycobacterium tuberculosis.
XX PT
PS Disclosure; Fig 4; 42pp; English.
XX PS
CC The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine
candidate peptides. The invention also relates to a method for
identifying Mtb vaccine candidate peptides as well as vaccines comprising
these candidate peptides. Vaccines of the invention and Mtb vaccine
candidate peptides are useful for inducing an anti-Mycobacterium
tuberculosis (anti-Mtb) immune response by raising anti-Mtb antibody in a
mammalian subject, preferably human. They are used for immunising a
mammalian subject, preferably humans, against infection caused by
Mycobacterium tuberculosis. The present sequence is a Mtb vaccine
candidate peptide
XX CC
SQ Sequence 20 AA;
Query Match 100.0%; Score 96; DB 4; Length 20;

26 45 46.9 266 6 ABM71199
27 45 46.9 870 2 AAR30729
28 45 46.9 1247 6 ABJ19352
29 44 45.8 263 6 ABU43727
30 43.5 45.3 1047 5 ABB93446
31 42.5 44.3 917 4 AAG70831
32 42 43.8 123 4 AAU17323
33 42 43.8 123 7 ABR94031
34 42 43.8 251 6 ABR58351
35 42 43.8 694 4 AAB93613
36 42 43.8 962 2 AAY31646
37 42 43.8 962 6 AAG79733
38 42 43.8 2548 2 AAY05781
39 41.5 43.2 2548 7 ADE55538
40 41 42.7 216 5 AAU84388
41 41 42.7 222 6 ABU27470
42 41 42.7 365 3 AAB11052
43 41 42.7 365 6 AAE36576
44 41 42.7 365 6 ABU61755
45 41 42.7 365 6 ABU61755

ABM71199 Staphyloc
AAR30729 p100 prot
ABJ19352 NOVX rela
ABU43727 Protein e
ABB93446 Herbicida
AAG70831 C albican
AAU17323 Novel sig
ABR94031 Human nov
ABR58351 Bcd215 pr
AAB93613 Human pro
AAY31646 Human tra
AAG79733 Human TAP
AAY05781 Human myo
ADE55538 Human pro
AAU84388 Novel hum
ABU27470 Protein e
AAB11052 A. naeslu
AAE36576 Actinomyc
ABU61755 Sortase-t

Best Local Similarity 100.0%; Pred. No. 7.2e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTYHVV 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1 IDELKTNSLLTSLTYHVV 20

RESULT 2

AAP91963
ID AAP91963 standard; protein; 163 AA.

XX AC AAP91963;

XX XX 25-MAR-2003 (revised)

DT DT 22-FEB-1990 (first entry)

XX DE MPB-70 protein of Mycobacterium bovis AN5.

XX XX Mycobacterium bovis; strain AN5; MPB-70 protein.

XX OS Mycobacterium bovis.

XX XX WO8909261-A.

XX PD 05-OCT-1989.

XX FF 31-MAR-1989; 89WO-AU000143.

XX PR 31-MAR-1988; 88AU-00007550.

XX XX (CSIR) COMMONWEALTH SCI & IND RES ORG.

XX PI Wood PR, Radford AJ;

XX DR WPI; 1989-309529/42.

XX DR N-PSDB; AAN91472.

XX XX Diagnosis of mycobacterium bovis infection - using antibodies specific to
MPB-70 protein of M bovis.

XX PS Disclosure; Fig 3; 48pp; English.

XX CC The MPB-70 protein can be detected by specific antibodies or by a cell-
mediated immune response against it, to diagnose M.bovis infection. The
protein is produced by chromatofocussing of a M.bovis AN5 culture
filtrate. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-
MAR-2003 to correct PA field.)

XX SQ Sequence 163 AA;

Query Match 100.0%; Score 96; DB 1; Length 163;

Best Local Similarity 100.0%; Pred. No. 9.5e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTYHVV 20
| | | | | | | | | | | | | | | | | | | | | |
Db 84 IDELKTNSLLTSLTYHVV 103

RESULT 3

AAW12045
ID AAW12045 standard; protein; 163 AA.

XX AC AAW12045;

XX DT 02-APR-1997 (first entry)

XX DE MPB 70.

XX XX T-cell epitope; antigen; T-cell determinant; receptor; MHC protein; bird;
KW HIV sf2; herpes simplex virus; antigen GD2; tetanus toxoid; vaccine; HSV;
KW mammal; gp120; immune response; B-cell antigen.

XX OS Mycobacterium bovis.

XX PN WO9518148-A1.

XX PD 06-JUL-1995.

XX PF 28-DEC-1993; 93WO-US011703.

XX PR 28-DEC-1993; 93WO-US011703.

XX PA (CHIR-) CHIRON MIMOTOPES PTY LTD.

XX XX Geysen HM, Rodda SJ;

XX DR WPI; 1995-246333/32.

XX PT T cell epitope peptide(s) - useful for detecting exposure of a subject to
an antigen or pathogen, and in vaccines for birds and mammals.

XX PS Example 1; Page 9-10; 57pp; English.

XX CC This sequence represents the Mycobacterium bovis MPB70. This sequence was
used to create a pool of T-cell epitope peptides (see AAW12045-W12067). T
-cell epitopes (also known as T-cell determinants) are peptides (or
regions of a protein) which bind to T-cell antigen receptors in
conjugation with MHC proteins. The epitope sequences shown in AAW11953-
W11976 were the most antigenic peptides obtained from pools of peptides
created from the HIV sf2 gp120 (AAW11953-W11960), herpes simplex virus
antigen GD2 (AAW11961-W11969), and tetanus toxoid (AAW11970-W11976). The
epitopes can be used in methods for detecting exposure of a mammal or
bird to an antigen, and for increasing the number of T-cells specific for
an antigen. The peptides can also be used in a method for determining T-
cell epitopes specific for an antigen. These methods allow for the
identification of T-cell determinants. The T-cell epitope peptides can be
used in a vaccine for inducing an immune response in a bird or mammal.
The vaccine also contains a B-cell antigen, preferably herpes simplex
virus GD2 (see AAW12068) or HIV sf2 gp120 (see AAW11977), and a carrier
Sequence 163 AA;

Query Match 100.0%; Score 96; DB 2; Length 163;

Best Local Similarity 100.0%; Pred. No. 9.5e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTYHVV 20

Db 84 IDELKTNSLLTSLTYHVV 103

RESULT 4

AAW07235

ID AAW07235 standard; protein; 192 AA.

XX AC AAW07235;

XX XX 25-MAR-2003 (revised)

DT DT 23-JAN-1991 (first entry)

XX DE Protein and and secretory region of MPB70 gene.

XX XX Actinomyces expression system; BCG; Mycobacteria.

XX OS Mycobacterium bovis.

XX PN WO9010701-A.

XX PD 20-SEP-1990.

XX PF 08-MAR-1989; 89AU-00003099.

XX PR 08-MAR-1989; 89AU-00003099.

XX XX

PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
 XX Radford A, Wood PR;
 PI
 XX WPI; 1990-305024/40.
 DR N-PSDB; AAO06112.
 XX
 XX DNA vectors contg. MPB70 gene promoter region - used as expression system
 PT for actinomycetes and related organisms, esp. mycobacterium bovis BCG.
 XX
 XX Disclosure; Fig 2; 31pp; English.
 XX
 XX Gene product was isolated from MPB70 gene of Mycobacterium bovis BCG.
 CC Product is encoded by a construct expressing heterologous antigenic genes
 CC for development of vaccines. Vaccines may be created which will induce a
 CC cell-mediated immune response, diagnostically distinguishable from wild
 CC type infection by Mycobacteria, eg. M.bovis, M.tuberculosis, M.lprae etc.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX
 XX Sequence 192 AA;
 SQ

Query Match 100.0%; Score 96; DB 2; Length 192;
 Best Local Similarity 100.0%; Pred. No. 1.2e-07;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHVW 20
 |||||
 DB 114 IDELKTNSLLTSILTYHVW 133

RESULT 5
 AAW99351
 ID AAW99351 standard; peptide; 193 AA.
 AC AAW99351;
 XX
 XX 21-MAY-1999 (first entry)
 DT
 XX M.tuberculosis MPT70 protein.
 DE
 XX MPT83; glycosylation motif; immunogenicity; vaccine; immunisation;
 KW mammal; infection; Mycobacterium tuberculosis; actinomycete.
 XX
 XX Mycobacterium tuberculosis.
 OS
 XX WO9902706-A1.
 PN
 XX 21-JAN-1999.
 PD
 XX 06-JUL-1998; 98WO-GB001989.
 PF
 XX 07-JUL-1997; 97GB-00014242.
 PR
 XX (UKAG-) UK MIN FISHERIES & FOOD.
 PA
 XX Hewinson RG, Michell SL;
 PI
 XX WPI; 1999-120907/10.
 DR
 XX A new recombinant DNA encoding a glycosylation motif - useful to
 PT glycosylate proteins when expressed in an actinomycete host.
 PF
 XX Disclosure; Page 17-18; 28pp; English.
 PS
 XX This sequence corresponds to the MPT70 protein from Mycobacterium
 CC tuberculosis. The invention relates to the introduction of protein
 CC glycosylation motifs, into other proteins, especially when expressed in
 CC an actinomycete host cell, in order to improve their immunogenicity and
 CC hence their use in e.g. vaccines. The polypeptide is used to immunise a
 CC mammal against infection by M. tuberculosis
 XX
 XX Sequence 193 AA;
 SQ

Query Match 100.0%; Score 96; DB 2; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1.2e-07;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHVW 20
 |||||
 DB 114 IDELKTNSLLTSILTYHVW 133

RESULT 6
 AAR06839
 ID AAR06839 standard; protein; 201 AA.
 XX
 XX AAR06839;
 AC
 XX 14-JAN-1991 (first entry)
 DT
 XX Human IL-2 N-terminal fragment and immunoprotein MPB 70 derived from a
 DE BCG bacteria.
 DE
 XX Bovine tuberculosis; Interleukin-2; IL-2; plasmid pT13S.
 KW
 XX Mycobacterium bovis.
 OS
 XX JP02195895-A.
 PN
 XX 02-AUG-1990.
 PD
 XX 24-JAN-1989; 89JP-00013270.
 PF
 XX 24-JAN-1989; 89JP-00013270.
 PR
 XX (AJIN) AJINOMOTO KK.
 PA
 XX WPI; 1990-278851/37.
 DR
 XX BCG bacteria derived immuno:protein MPB70 - can be used as diagnostic
 PT agent used to determine bovine tuberculosis.
 PT
 XX Disclosure; Fig 3; 11pp; Japanese.
 PS
 XX Immunoprotein MPB 70 encoding sequence may be incorporated into plasmid
 CC pT13S with an N-terminal fragment of human IL-2. The plasmid may be used
 CC to transform an expression system giving a fusion protein which may be
 CC used as a diagnostic agent for bovine tuberculosis infection
 CC
 XX Sequence 201 AA;
 SQ

Query Match 100.0%; Score 96; DB 2; Length 201;
 Best Local Similarity 100.0%; Pred. No. 1.2e-07;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHVW 20
 |||||
 DB 122 IDELKTNSLLTSILTYHVW 141

RESULT 7
 AAR07053
 ID AAR07053 standard; protein; 163 AA.
 XX
 XX AAR07053;
 AC
 XX 14-JAN-1991 (first entry)
 DT
 XX Immunoprotein MPB 70 derived from a BCG bacteria.
 DE
 XX Bovine tuberculosis; Interleukin-2; IL-2; plasmid pT13S.
 KW
 XX Mycobacterium bovis.
 OS
 XX JP02195895-A.
 PN
 XX
 XX Sequence 163 AA;
 SQ

PD 02-AUG-1990.
 XX 24-JAN-1989; 89JP-00013270.
 PF 24-JAN-1989; 89JP-00013270.
 PR (AJIN) AJINOMOTO KK.
 XX WPI; 1990-278851/37.
 XX N-PSDB; AAQ05975.
 XX BCG bacteria derived immuno:protein MPB70 - can be used as diagnostic
 PT agent used to determine bovine tuberculosis.
 PT Claim 1; Page 694; 11pp; Japanese.
 XX Immunoprotein MPB 70 encoding sequence may be incorporated into plasmid
 CC pfl3S with an N-terminal fragment of human IL-2. The plasmid may be used
 CC to transform an expression system giving a fusion protein which may be
 CC used as a diagnostic agent for bovine tuberculosis infection
 XX Sequence 163 AA;
 SQ

Query Match 95.8%; Score 92; DB 2; Length 163;
 Best Local Similarity 95.0%; Pred. No. 4.4e-07;
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 IDELKTNSLLTSILTYHV 20
 DB 84 IEELKTNSLLTSILTYHV 103

RESULT 8
 AAW17979
 ID AAW17979 standard; protein; 220 AA.
 XX AAW17979;
 DT 23-JUL-1997 (first entry)
 XX Mycobacterium tuberculosis mpt83 protein.
 DE Vaccine; diagnostic agent; antigen; lipoylation; glycosylation.
 XX Mycobacterium tuberculosis.
 XX Key Location/Qualifiers
 FH Peptide 1..21
 FT /label= Secretion_signal
 FT Misc-difference 22..26
 FT /label= Lipoylation_motif
 FT Misc-difference 35..57
 FT /label= Glycosylation_motif
 XX WO9708322-A1.
 PN 06-MAR-1997.
 XX 19-AUG-1996; 96WO-GB002015.
 XX 25-AUG-1995; 95GB-00017494.
 XX (UKAG-) UK MIN FISHERIES & FOOD.
 XX Hewinson RG, Jacobs WR;
 XX WPI; 1997-179279/16.
 XX N-PSDB; AAT70155.
 XX New isolated mpt83 gene from Mycobacterium tuberculosis - used to develop
 PT prods. for use as vaccines or as diagnostic agents.
 XX Claim 1; Page 23-24; 40pp; English.

XX The gene mpt83 encodes the 25 kDa antigen of Mycobacterium tuberculosis.
 CC The present sequence represents a 220 amino acid mpt83 gene product. The
 CC protein can be used in vaccines to protect against Mycobacterium
 CC tuberculosis infection or for antibody production. The antibodies are
 CC useful as diagnostic agents, to detect M. tuberculosis infection in a
 CC sample. The mpt83 gene promoter, glycosylation, lipoylation and secretion
 CC sequences can be used in recombinant DNA expression systems for use in
 CC the transformation of cells e.g. to produce glycosylated or lipoylated
 CC products which can be regulated by macrophage factors
 XX Sequence 220 AA;
 SQ

Query Match 79.2%; Score 76; DB 2; Length 220;
 Best Local Similarity 70.0%; Pred. No. 0.00028;
 Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 IDELKTNSLLTSILTYHV 20
 DB 140 IDQLKTDKLLSSILTYHVI 159

RESULT 9
 AAW99350
 ID AAW99350 standard; peptide; 220 AA.
 XX AAW99350;
 DT 21-MAY-1999 (first entry)
 XX M.tuberculosis MPT83 protein.
 DE MPT83; Glycosylation motif; immunogenicity; vaccine; immunisation;
 KW mammal; infection; Mycobacterium tuberculosis; actinomycete.
 XX Mycobacterium tuberculosis.
 XX WO9902706-A1.
 PN 21-JAN-1999.
 XX 06-JUL-1998; 98WO-GB001989.
 XX 07-JUL-1997; 97GB-00014242.
 XX (UKAG-) UK MIN FISHERIES & FOOD.
 XX Hewinson RG, Michell SL;
 XX WPI; 1999-120907/10.
 XX A new recombinant DNA encoding a glycosylation motif - useful to
 XX glycosylate proteins when expressed in an actinomycete host.
 XX Disclosure; Page 16-17; 28pp; English.
 XX This sequence corresponds to the MPT83 protein from Mycobacterium
 CC tuberculosis. The invention relates to the introduction of protein
 CC glycosylation motifs, into other proteins, especially when expressed in
 CC an actinomycete host cell, in order to improve their immunogenicity and
 CC hence their use in e.g. vaccines. The polypeptide is used to immunise a
 CC mammal against infection by M. tuberculosis
 XX Sequence 220 AA;
 SQ

Query Match 79.2%; Score 76; DB 2; Length 220;
 Best Local Similarity 70.0%; Pred. No. 0.00028;
 Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 IDELKTNSLLTSILTYHV 20
 DB 140 IDQLKTDKLLSSILTYHVI 159

RESULT 11


```

PF 02-JUN-1999; 99US-00324542.
XX
PR 23-DEC-1997; 97US-00997080.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Watson JD, Tan PLJ, Prestidge R;
XX
DR WPI; 2002-138361/18.
XX
DR N-PSDB; ABL36269.
XX
PT Inhibiting skin inflammation associated with skin disorder e.g.
PT psoriasis, by administering composition comprising deslipidated and
PT deglycolipidated Mycobacterium vaccae cells or Mycobacterium vaccae
PT culture filtrate.
XX
XX
PS Example 5; Col 153-156; 116pp; English.
XX
CC The present invention relates to a method of inhibiting skin inflammation
CC associated with a skin disorder selected from psoriasis, atopic
CC dermatitis and allergic contact dermatitis, which involves administering
CC a composition containing deslipidated and deglycolipidated Mycobacterium
CC vaccae cells or M. vaccae culture filtrate. The skin disorder to be
CC treated may also include alopecia areata, and skin cancers such as basal
CC cell carcinoma, squamous cell carcinoma and melanoma. The composition
CC acts by inhibiting the Th2 immune response. The present sequence is a
CC protein described in the exemplification of the invention
XX
XX Sequence 228 AA;
SQ
Query Match 72.9%; Score 70; DB 5; Length 228;
Best Local Similarity 65.0%; Pred. No. 0.0029;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHV 20
Db 148 LETLKTSDMLTNILTYHV 167

Search completed: March 10, 2004, 12:05:38
Job time : 49.5641 secs

PT and Toll-like receptor signaling pathways, and treating e.g. autoimmune
PT disorders.
XX
XX Claim 7; Page 114-115; 136pp; English.
XX
CC The present invention relates to methods for modulating immune responses
CC by modulating the Notch signalling and Toll-like receptor signalling
CC pathways using compositions comprising mycobacteria antigens (ACC42518-
CC ACC42543 and ABP70878-ABP70903). The methods are useful for modulating
CC immune responses and treating e.g. autoimmune disorders (such as multiple
CC sclerosis, rheumatoid arthritis, Type 1 diabetes mellitus, psoriasis,
CC systemic lupus erythematosus, scleroderma), allergic disease and graft
CC rejection and also disorders characterised by undesired apoptotic cell
CC death or undesired cell proliferation
XX
XX Sequence 228 AA;
SQ
Query Match 72.9%; Score 70; DB 6; Length 228;
Best Local Similarity 65.0%; Pred. No. 0.0029;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHV 20
Db 148 LETLKTSDMLTNILTYHV 167

Search completed: March 10, 2004, 12:05:38
Job time : 49.5641 secs

PF 02-JUN-1999; 99US-00324542.
XX
PR 23-DEC-1997; 97US-00997080.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Watson JD, Tan PLJ, Prestidge R;
XX
DR WPI; 2002-138361/18.
XX
DR N-PSDB; ABL36269.
XX
PT Inhibiting skin inflammation associated with skin disorder e.g.
PT psoriasis, by administering composition comprising deslipidated and
PT deglycolipidated Mycobacterium vaccae cells or Mycobacterium vaccae
PT culture filtrate.
XX
XX
PS Example 5; Col 153-156; 116pp; English.
XX
CC The present invention relates to a method of inhibiting skin inflammation
CC associated with a skin disorder selected from psoriasis, atopic
CC dermatitis and allergic contact dermatitis, which involves administering
CC a composition containing deslipidated and deglycolipidated Mycobacterium
CC vaccae cells or M. vaccae culture filtrate. The skin disorder to be
CC treated may also include alopecia areata, and skin cancers such as basal
CC cell carcinoma, squamous cell carcinoma and melanoma. The composition
CC acts by inhibiting the Th2 immune response. The present sequence is a
CC protein described in the exemplification of the invention
XX
XX Sequence 228 AA;
SQ
Query Match 72.9%; Score 70; DB 5; Length 228;
Best Local Similarity 65.0%; Pred. No. 0.0029;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHV 20
Db 148 LETLKTSDMLTNILTYHV 167

RESULT 15
ABP70879
ID ABP70879 standard; protein; 228 AA.
XX
AC ABP70879;
XX
DT 26-AUG-2003 (first entry)
XX
DE Mycobacterium vaccae antigen GV-1/83, SEQ ID 28.
XX
KW Immunosuppressive; neuroprotective; antirheumatic; antiarthritic;
KW antidiabetic; antipsoriatic; dermatological; anti-inflammatory;
KW immune response; Notch signalling pathway; autoimmune disorder;
KW Toll-like receptor signalling pathway; antigen; allergy; graft rejection;
KW apoptotic cell death; cell proliferation.
XX
OS Mycobacterium vaccae.
XX
XX WO2003013595-A1.
XX
XX 20-FEB-2003.
XX
XX 26-JUL-2002; 2002WO-NZ000135.
XX
XX 26-JUL-2001; 2001US-0308446P.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Watson JD, Tan PLJ, Abernethy N;
XX
DR WPI; 2003-239567/23.
XX
DR N-PSDB; ACC42519.
XX
PT Methods for modulating immune responses by modulating the Notch signaling

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 12:11:07 ; Search time 25.1282 Seconds
(without alignments)
168.061 Million cell updates/sec

Title: US-10-044-703-78

Perfect score: 96
Sequence: 1 IDELKTNSLLTSILTYHV 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	20	9	US-09-813-333-78
2	96	100.0	20	13	US-10-044-703-78
3	71	74.0	231	10	US-09-880-505-152
4	71	74.0	231	13	US-10-051-643-152
5	71	74.0	231	14	US-10-205-379-27
6	70	72.9	228	10	US-09-880-505-147
7	70	72.9	228	13	US-10-051-643-147
8	70	72.9	228	14	US-10-205-379-28
9	50	52.1	149	15	US-10-028-248A-44
10	50	52.1	149	15	US-10-028-248A-45
11	50	52.1	149	15	US-10-107-782-44
12	50	52.1	149	15	US-10-107-782-45
13	49	51.0	20	9	US-09-813-333-79
14	49	51.0	20	13	US-10-044-703-79
15	48	50.0	215	14	US-10-156-761-13069

16	47	49.0	20	9	US-09-813-333-77	Sequence 77, Appl
17	47	49.0	20	13	US-10-044-703-77	Sequence 77, Appl
18	46	47.9	180	13	US-10-047-280-18	Sequence 18, Appl
19	45	46.9	281	9	US-09-815-242-5376	Sequence 5376, Ap
20	45	46.9	266	9	US-09-815-242-12291	Sequence 12291, A
21	45	46.9	1247	15	US-10-161-493-70	Sequence 70, Appl
22	42	43.8	123	9	US-09-764-868-888	Sequence 888, App
23	42	43.8	327	15	US-10-369-493-2729	Sequence 2729, Ap
24	42	43.8	667	14	US-10-394-136-55	Sequence 55, Appl
25	42	43.8	2548	9	US-09-851-682A-1	Sequence 1, Appl
26	41.5	43.2	846	9	US-09-376-045-2	Sequence 2, Appl
27	41.5	43.2	908	9	US-09-376-045-6	Sequence 6, Appl
28	41.5	43.2	1053	14	US-10-217-700-6	Sequence 6, Appl
29	41.5	43.2	1054	9	US-09-376-045-4	Sequence 4, Appl
30	41	42.7	365	10	US-09-933-999A-5	Sequence 5, Appl
31	41	42.7	365	14	US-10-219-700-5	Sequence 5, Appl
32	41	42.7	658	14	US-10-288-757-87	Sequence 87, Appl
33	41	42.7	670	14	US-10-289-757-176	Sequence 176, App
34	41	42.7	670	16	US-10-389-566-1502	Sequence 182, Ap
35	40	41.7	47	14	US-10-144-929-182	Sequence 182, App
36	40	41.7	47	15	US-10-144-929-182	Sequence 182, App
37	40	41.7	56	14	US-10-106-698-6847	Sequence 6847, AP
38	40	41.7	568	15	US-10-369-493-10890	Sequence 10890, A
39	40	41.7	683	14	US-10-081-872-188	Sequence 188, App
40	40	41.7	683	15	US-10-385-305-188	Sequence 188, App
41	39.5	41.1	1074	14	US-10-289-757-72	Sequence 72, Appl
42	39.5	41.1	1075	14	US-10-289-757-167	Sequence 167, App
43	39	40.6	272	9	US-09-815-243-5738	Sequence 5738, Ap
44	39	40.6	281	9	US-09-815-243-12170	Sequence 12170, A
45	39	40.6	281	9	US-09-815-242-13016	Sequence 13016, A

ALIGNMENTS

RESULT 1

US-09-813-333-78
; Sequence 78, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:
; APPLICANT: Decroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-78

Query Match 100.0%; Score 96; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	IDELKTNSLLTSILTYHV 20
Db	1	IDELKTNSLLTSILTYHV 20

RESULT 2

US-10-044-703-78
; Sequence 78, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:
; APPLICANT: DeGroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703


```
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 78
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-78

Query Match          100.0%; Score 96; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHV 20
Db 1 IDELKTNSLLTSILTYHV 20

RESULT 3
US-09-880-505-152
; Sequence 152, Application US/09880505
; Publication No. US20030007976A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L.J.
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
; FILE REFERENCE: 11000.1007c2
; CURRENT APPLICATION NUMBER: US/09/880,505
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 09/324,542
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: US 08/997,080
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 152
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-09-880-505-152

Query Match          74.0%; Score 71; DB 10; Length 231;
Best Local Similarity 70.0%; Pred. No. 0.0014;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHV 20
Db 136 LETLKTDSDLLTKILTYHV 155

RESULT 4
US-10-051-643-152
; Sequence 152, Application US/10051643
; Publication No. US20020197265A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L.J.
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Diseases of the Respiratory
; TITLE OF INVENTION: System using Mycobacterium Vaccae
; FILE REFERENCE: 11000.1008c2
; CURRENT APPLICATION NUMBER: US/10/051,643
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US09/156,181
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: US 08/996,624
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 152
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-10-051-643-152

Query Match          74.0%; Score 71; DB 13; Length 231;
Best Local Similarity 70.0%; Pred. No. 0.0014;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHV 20
Db 136 LETLKTDSDLLTKILTYHV 155

RESULT 5
US-10-205-979-27
; Sequence 27, Application US/10205979
; Publication No. US20030147861A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L.J.
; APPLICANT: Abernethy, Nevin
; TITLE OF INVENTION: Compounds and Methods for the Modulation
; TITLE OF INVENTION: of Immune Responses
; FILE REFERENCE: 11000.1063U
; CURRENT APPLICATION NUMBER: US/10/205,979
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/308,446
; PRIOR FILING DATE: 2001-07-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-10-205-979-27

Query Match          74.0%; Score 71; DB 14; Length 231;
Best Local Similarity 70.0%; Pred. No. 0.0014;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHV 20
Db 136 LETLKTDSDLLTKILTYHV 155

RESULT 6
US-09-880-505-147
; Sequence 147, Application US/09880505
; Publication No. US20030007976A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L.J.
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
; FILE REFERENCE: 11000.1007c2
; CURRENT APPLICATION NUMBER: US/09/880,505
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 09/324,542
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: US 08/997,080
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 147
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-09-880-505-147

Query Match          72.9%; Score 70; DB 10; Length 228;
```

```

QY      1 IDELKTNSLLTSILTYHV 20
      : : : : : : : : : :
DB     148 LETLKTDSMLNLTLYHV 167

```

RESULT 10
US-10-028-248A-45
; Sequence 45, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shinkets, Richard

```
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malyankar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr., Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Brenda
; APPLICANT: Zernhusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: No. US2003023582A1el Nucleic Acids and Polypeptides and Methods
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fasciclin
US-10-028-248A-45

Query Match 52.1%; Score 50; DB 15; Length 149;
Best Local Similarity 54.5%; Pred. No. 2.3;
Matches 12; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

Qy 1 IDEL--KTNSLLTSLTYHV 20
Db 57 LDELLNKEDAKQLKILTYHV 78

RESULT 11
US-10-107-782-44
; Sequence 44, Application US/10107782
; Publication No. US20040018970A1
; GENERAL INFORMATION:
; APPLICANT: Boldog, Ferenc
; APPLICANT: Casman, Stacie
; APPLICANT: Colman, Steve
; APPLICANT: Edinger, Shlomit
; APPLICANT: Gangolli, Esha
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Malyankar, Uriel
; APPLICANT: Miller, Charles
; APPLICANT: Patturajan, Meera
; APPLICANT: Rothenberg, Mark
; APPLICANT: Sciore, Paul
; APPLICANT: Shenoy, Suresh
```

```
; APPLICANT: Malyankar, Uriel
; APPLICANT: Miller, Charles
; APPLICANT: Millet, Isabelle
; APPLICANT: Patturajan, Meera
; APPLICANT: Rothenberg, Mark
; APPLICANT: Sciore, Paul
; APPLICANT: Shenoy, Suresh
; APPLICANT: Shimkets, Richard
; APPLICANT: Si, Jingsheng
; APPLICANT: Smithson, Glennda
; APPLICANT: Spytek, Kimberly
; APPLICANT: Stone, David
; APPLICANT: Taupier, Raymond, Jr.,
; APPLICANT: Tchernev, Velizar
; APPLICANT: Vernet, Corine
; APPLICANT: Zernhusen, Brian
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 21402-222CIP
; CURRENT APPLICATION NUMBER: US/10/107,782
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 10/028,248
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256,619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262,959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272,408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285,189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308,039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311,266
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/279,344
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: GuraSeqList version 0.1
; SEQ ID NO 44
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-107-782-44

Query Match 52.1%; Score 50; DB 15; Length 149;
Best Local Similarity 54.5%; Pred. No. 2.3;
Matches 12; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

Qy 1 IDEL--KTNSLLTSLTYHV 20
Db 57 LDELLNKEDAKQLKILTYHV 78

RESULT 12
US-10-107-782-45
; Sequence 45, Application US/10107782
; Publication No. US20040018970A1
; GENERAL INFORMATION:
; APPLICANT: Boldog, Ferenc
; APPLICANT: Casman, Stacie
; APPLICANT: Colman, Steve
; APPLICANT: Edinger, Shlomit
; APPLICANT: Gangolli, Esha
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Malyankar, Uriel
; APPLICANT: Miller, Charles
; APPLICANT: Millet, Isabelle
; APPLICANT: Patturajan, Meera
; APPLICANT: Rothenberg, Mark
; APPLICANT: Sciore, Paul
; APPLICANT: Shenoy, Suresh
```

```
; APPLICANT: Shimkets, Richard,
; APPLICANT: Si, Jingsheng,
; APPLICANT: Smithson, Glennda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Stone, David,
; APPLICANT: Taupier, Raymond, jr.,
; APPLICANT: Tchernev, Velizar,
; APPLICANT: Vernet, Corine,
; APPLICANT: Zernhusen, Brian
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 21402-222C1P
; CURRENT APPLICATION NUMBER: US/10/107,782
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 10/028,248
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256,619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262,959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272,408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285,189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308,039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311,266
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/279,344
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 45
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-107-782-45
```

```
Query Match 52.1%; Score 50; DB 15; Length 149;
Best Local Similarity 54.5%; Pred. No. 2.3;
Matches 12; Conservative 3; Mismatches 5; Indels 2; Gaps 1;
```

```
Qy 1 IDEL--KTNSSLTSLTYHV 20
Db 57 LDELLNKEDAKQLALTYHV 78
```

```
RESULT 13
US-09-813-333-79
; Sequence 79, Application US/09813333
; Patent No. US20020119160A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 79
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-73
```

```
Query Match 51.0%; Score 49; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 11 LTSILTYHV 20
Db 1 LTSILTYHV 10
```

```
RESULT 14
US-10-044-703-79
; Sequence 79, Application US/10044703
; Publication No. US2002019233A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 79
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-73
```

```
Query Match 51.0%; Score 49; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 11 LTSILTYHV 20
Db 1 LTSILTYHV 10
```

```
RESULT 15
US-10-156-761-13069
; Sequence 13069, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13069
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13069
```

```
Query Match 50.0%; Score 48; DB 14; Length 215;
Best Local Similarity 42.1%; Pred. No. 7.3;
Matches 8; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 1 IDELKTNSSLTSLTYHV 19
Db 110 LDHVKTNTVNTAMTLTAHL 128
```

```
Search completed: March 10, 2004, 12:41:32
Job time : 25.1282 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:58:01 ; Search time 10.7692 Seconds
(without alignments)
178.641 Million cell updates/sec

Title: US-10-044-703-78

Perfect score: 96
Sequence: 1 IDELKTNSLLTSILTYHV 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78:*

1: Pir1:*

2: Pir2:*

3: Pir3:*

4: Pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96	100.0	193	2 A48320	major secreted imm
2	96	100.0	193	2 A37195	major secreted pro
3	96	100.0	193	2 F70923	probable mpt70 pro
4	80	83.3	84	2 A43502	MBP70 protein - My
5	76	73.2	220	2 D70923	probable mpt83 pro
6	55	57.3	3255	2 G81702	adherence factor T
7	49	51.0	133	2 A51971	hypothetical prote
8	49	51.0	160	2 F95334	Nex18 Symbiotical
9	47	49.0	623	2 F75523	osteoblast specifi
10	46	47.9	180	2 S76811	transforming growt
11	46	47.9	180	2 G71872	transforming growt
12	45	46.9	261	2 A2280	hypothetical prote
13	45	46.9	263	2 A99960	conserved hypotet
14	45	46.9	871	1 XPBE12	major antigenic st
15	44	44.8	141	2 AF2417	hypothetical prote
16	44	44.8	220	2 T34986	probable secreted
17	43.5	45.3	1047	2 T51800	sucrose-phosphate
18	43	44.8	133	2 S77329	secreted protein M
19	43	44.8	255	2 B97861	hypothetical prote
20	43	44.8	274	1 S1872	rolB protein - Agr
21	43	44.8	573	2 E69802	ABC transporter (A
22	43	44.8	894	2 T13029	beta-adaptin homol
23	42	43.8	150	2 T28731	hypothetical prote
24	42	43.8	327	2 S75506	porphobilinogen sy
25	42	43.8	356	2 T33918	hypothetical prote
26	42	43.8	358	2 T15369	hypothetical prote
27	42	43.8	403	2 T02290	hypothetical prote
28	42	43.8	452	2 B59096	hypothetical prote
29	42	43.8	558	2 AE2033	hypothetical prote

ALIGNMENTS

RESULT 1

A48320
major secreted immunogenic protein MPB70 precursor - Mycobacterium bovis (strain BCG)
C/Species: Mycobacterium bovis
C/Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 10-Mar-1994
C/Accession: A48320

R/Terasaka, K.; Yamaguchi, R.; Matsuo, K.; Yamazaki, A.; Nagai, S.; Yamada, T.
FEMS Microbiol. Lett. 58, 273-276, 1989
A/Title: Complete nucleotide sequence of immunogenic protein MPB70 from Mycobacterium

A/Reference number: A48320
A/Accession: A48320
A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-193 <TER>

A/Cross-references: GB:X17086

Query Match 100.0%; Score 96; DB 2; Length 193;

Best Local Similarity 100.0%; Pred. No. 2.7e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IDELKTNSLLTSILTYHV 20

Db 114 IDELKTNSLLTSILTYHV 133

RESULT 2

A37195
major secreted protein MPB70 precursor - Mycobacterium bovis

C/Species: Mycobacterium bovis

C/Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 22-Oct-1999

C/Accession: A37195

R/Radford, A.J.; Wood, P.R.; Billman-Jacobe, H.; Geysen, H.M.; Mason, T.J.; Tribbick, G.

J. Gen. Microbiol. 136, 265-272, 1990

A/Title: Epitope mapping of the Mycobacterium bovis secretory protein MPB70 using overl

A/Reference number: A37195; MUID:90218009; PMID:1691265

A/Accession: A37195

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-193 <RAD>

A/Cross-references: GB:M33916; GB:M25396; NID:G149975; PIDN:AAA25366.1; PID:G149976

Query Match 100.0%; Score 96; DB 2; Length 193;

Best Local Similarity 100.0%; Pred. No. 2.7e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IDELKTNSLLTSILTYHV 20

Db 114 IDELKTNSLLTSILTYHV 133

RESULT 3

F70923

probable mpt70 protein - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: F70923
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: F70923
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-193 <COL>
 A: Cross-references: GB:Z74024; GB:AL123456; NID:G3250700; PIDN:CAA98373.1; PID:e354630;
 A: Experimental source: strain H37RV
 C: Genetix:
 A: Gene: mpt70

Query Match 100.0%; Score 96; DB 2; Length 193;
 Best Local Similarity 100.0%; Pred. No. 2.7e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTYHV 20
 |||||
 DB 114 IDELKTNSLLTSLTYHV 133
 |||||

RESULT 4
 A43502
 MPB70 protein - Mycobacterium bovis (fragment)
 C: Species: Mycobacterium bovis
 C: Date: 21-Oct-1992 #sequence_revision 21-Oct-1992 #text_change 23-Jun-1993
 C: Accession: A43502
 R: Radford, A.J.; Duffield, B.J.; Plackett, P.
 Infect. Immun. 56, 921-925, 1988
 A: Title: Cloning of a species-specific antigen of Mycobacterium bovis.
 A: Reference number: A43502; MUID: 88153076; PMID: 3278986
 A: Accession: A43502
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-84 <RAD>
 A: Cross-references: GB:M7840

Query Match 83.3%; Score 80; DB 2; Length 84;
 Best Local Similarity 100.0%; Pred. No. 4.7e-06;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTY 17
 |||||
 DB 68 IDELKTNSLLTSLTY 84
 |||||

RESULT 5
 D70923
 probable mpt83 protein - Mycobacterium tuberculosis (strain H37RV)
 C: Species: Mycobacterium tuberculosis
 C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C: Accession: D70923
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: D70923
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-220 <COL>
 A: Cross-references: GB:Z74024; GB:AL123456; NID:G3250700; PIDN:CAA98350.1; PID:e1301024;
 A: Experimental source: strain H37RV

C: Genetix:
 A: Gene: mpt83

Query Match 79.2%; Score 76; DB 2; Length 220;
 Best Local Similarity 70.0%; Pred. No. 6.2e-05;
 Matches 14; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTYHV 20
 |||||
 DB 140 IDQLKTDAKLLSSILTYHV 159
 |||||

RESULT 6
 G81702
 adherence factor TC0437 [imported] - Chlamydia muridarum (strain Nigg)
 C: Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
 C: Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
 C: Accession: G81702
 R: Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A: Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
 A: Reference number: A81500; MUID: 20150255; PMID: 10684935
 A: Accession: G81702
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-3255 <RET>
 A: Cross-references: GB:AE002311; GB:AE002160; NID:G7190477; PIDN:AAF39291.1; PID:G719048
 A: Experimental source: strain Nigg (MoPn)
 C: Genetix:
 A: Gene: TC0437

Query Match 57.3%; Score 55; DB 2; Length 3255;
 Best Local Similarity 50.0%; Pred. No. 3.6;
 Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTYHV 20
 |||||
 DB 1711 LDEKLNGLVDLTDFLTKHVL 1730
 |||||

RESULT 7
 AE1971
 hypothetical protein alr1320 [imported] - Nostoc sp. (strain PCC 7120)
 C: Species: Nostoc sp. PCC 7120
 A: Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C: Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C: Accession: AE1971
 R: Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A: Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A: Reference number: AB1807; MUID: 21595285; PMID: 11759840
 A: Accession: AE1971
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-133 <KUR>
 A: Cross-references: GB:BA000019; PIDN:BA073277.1; PID:G17130667; GSPDB:GNC00179
 A: Experimental source: strain PCC 7120
 C: Genetix:
 A: Gene: alr1320

Query Match 51.0%; Score 49; DB 2; Length 133;
 Best Local Similarity 60.0%; Pred. No. 1.1;
 Matches 12; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTYHV 20
 |||||
 DB 54 IQTLVONIPQLTRILTYHV 73
 |||||

RESULT 8
 F95334

Nex18 Symbiotically induced conserved protein nex18 [imported] - Sinorhizobium meliloti
 C/Species: Sinorhizobium meliloti
 C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C/Accession: F95334
 R/Barnett, M.J.; Fisher, R.H.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowne, K.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surryski, R.; Wells, D.H.; Yeh, K.C.; Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
 A/Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
 A/Reference number: A95262; MUID:21396509; PMID:11481432
 A/Accession: F95334
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-160 <KUR>
 A/Cross-references: GB:AE006469; PIDN:AAK65240.1; PID:G14523630; GSPDB:GN00165
 A/Experimental source: strain 1021, megaplasmid pSynA
 R/Gallbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T. Science 293, 668-672, 2001
 A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, P.; Vandenbol, M.; Vorholter, P.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.C.
 A/Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A/Reference number: A96039; MUID:21368234; PMID:11474104
 A/Contents: annotation
 C/Genetics:
 A/Gene: nex18
 A/Genome: plasmid

Query Match 51.0%; Score 49; DB 2; Length 160;
 Best Local Similarity 71.4%; Pred. No. 1.3;
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 7 NSSLLTSILTYHV 20
 DB 84 NKQKLTILTYHV 97
 RESULT 9
 F75523
 osteoblast specific factor 2-related protein - Deinococcus radiodurans (strain R1)
 C/Species: Deinococcus radiodurans
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
 C/Accession: F75523
 R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M. Science 286, 1571-1577, 1999
 A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A/Reference number: A75250; MUID:20036896; PMID:10567266
 A/Accession: F75523
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-623 <WHI>
 A/Cross-references: GB:AE001900; GB:AE000513; NID:G6458079; PIDN:AAF09979.1; PID:G645808
 A/Experimental source: strain R1
 C/Genetics:
 A/Gene: DR0399
 A/Map position: 1

Query Match 49.0%; Score 47; DB 2; Length 623;
 Best Local Similarity 40.0%; Pred. No. 12;
 Matches 8; Conservative 7; Mismatches 5; Indels 0; Gaps 0;
 QY 1 IDELKTNSLLTSILTYHV 20
 DB 538 LNAVADPALLKQVLSYHV 557
 RESULT 10
 S76811
 transforming growth factor-induced protein - Synecchocystis sp. (strain PCC 6803)
 N/Alternate names: protein all1483
 C/Species: Synecchocystis sp.

A/Variety: PCC 6803
 C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C/Accession: S76811
 R/Kaneko, T.; Sato, S.; Korani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O.K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuc DNA Res. 3, 109-136, 1996
 A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis.
 A/Reference number: S74322; MUID:97061201; PMID:8905231
 A/Accession: S76811
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-180 <KAN>
 A/Cross-references: EMBL:D90916; GB:AB001339; NID:G1653715; PIDN:BAAL8723.1; PID:G10194
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 47.9%; Score 46; DB 2; Length 180;
 Best Local Similarity 58.8%; Pred. No. 4.6;
 Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 4 LKTNSSLLTSILTYHV 20
 DB 103 LPENKDLVKILTYHV 119

RESULT 11
 G71872
 2,3-bisphosphoglycerate-independent phosphoglycerate mutase - Helicobacter pylori (stra
 C/Species: Helicobacter pylori
 A/Variety: strain J99
 C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 27-Oct-2003
 C/Accession: G71872
 R/Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999
 A/Title: Genomic sequence comparison of two unrelated isolates of the human gastric pat
 A/Reference number: A71800; MUID:99120557; PMID:9923682
 A/Accession: G71872
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-491 <ARN>
 A/Cross-references: GB:AB001520; GB:AE001439; NID:G4155483; PIDN:AAD06490.1; PID:G41554
 A/Experimental source: strain J99
 C/Genetics:
 A/Gene: pgm
 C/Superfamily: cofactor-independent phosphoglycerate mutase

Query Match 47.9%; Score 46; DB 2; Length 491;
 Best Local Similarity 52.6%; Pred. No. 14;
 Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 2 DELKTNSSLLTSILTYHV 20
 DB 88 DELKNNPFLNTIQSHV 106

RESULT 12
 AF2280
 hypothetical protein all3797 [imported] - Nostoc sp. (strain PCC 7120)
 C/Species: Nostoc sp. PCC 7120
 A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C/Accession: AF2280
 R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, DNA Res. 8, 205-213, 2001
 A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
 A/Reference number: AB1807; MUID:21595285; PMID:11759840
 A/Accession: AF2280
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-261 <KUR>
 A/Cross-references: GB:BA000019; PIDN:BA075496.1; PID:G17132931; GSPDB:GN00179

A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all3797

Query Match 46.9%; Score 45; DB 2; Length 261;
 Best Local Similarity 57.1%; Pred. No. 10;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 NSSLTSLTYHV 20
 Db 186 NKEVLKVLTYHV 199

RESULT 13

A89960
 conserved hypothetical protein SA1570 [imported] - Staphylococcus aureus (strain N315)
 C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: A89960
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayaashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: A89960
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-263 <KUR>
 A:Cross-references: GB:BA000018; PID:g13701544; PIDN:BA842838.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: SA1570
 C:Superfamily: Bacillus subtilis hypothetical protein ytmP

Query Match 46.9%; Score 45; DB 2; Length 263;
 Best Local Similarity 50.0%; Pred. No. 10;
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 KTNSSLLTSLTYHV 20
 Db 119 KINASLREVLTHIV 134

RESULT 14

XPBE12
 major antigenic structural protein p100 - human herpesvirus 6 (strain UI102)
 C:Species: human herpesvirus 6
 C:Date: 30-Jun-1993 #sequence_revision 29-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T09303; A42533
 R:Nicholas, J.; Martin, M.
 J. Virol. 68, 597-610, 1994
 A:Title: Nucleotide sequence analysis of a 38.5-kilobase-pair region of the genome of hu
 A:Reference number: Z16644; MUID:94118404; PMID:8289364
 A:Accession: T09303
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-871 <NIC>
 A:Cross-references: EMBL:L25528; NID:G451932; PIDN:AAA16716.1; PID:G451934
 R:Neipel, F.; Ellinger, K.; Fleckenstein, B.
 J. Virol. 66, 3918-3924, 1992
 A:Title: Gene for the major antigenic structural protein (p100) of human herpesvirus 6.
 A:Reference number: A42533; MUID:92260671; PMID:1374813
 A:Accession: A42533
 A:Molecule type: DNA
 A:Residues: 2-871 <NEI>
 A:Cross-references: GB:M87287; NID:G330673; PIDN:AAA46012.1; PID:G330674
 C:Genetics:
 A:Gene: P10F1
 C:Superfamily: human herpesvirus large structural phosphoprotein; large structural phosph
 C:Keywords: phosphoprotein
 F, 7-368/Domain: large structural phosphoprotein homology <CLS>

Query Match 46.9%; Score 45; DB 1; Length 871;
 Best Local Similarity 50.0%; Pred. No. 38;
 Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTYHV 20
 Db 55 VDDLKTLNLLVLMYHV 74

RESULT 15

AF2417
 hypothetical protein all4994 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AF2417
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AF2417
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-141 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA876593.1; PID:g17134032; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all4894

Query Match 45.8%; Score 44; DB 2; Length 141;
 Best Local Similarity 50.0%; Pred. No. 7.6;
 Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSLTYHV 20
 Db 62 VDALLKDIPKXILTYHV 81

Search completed: March 10, 2004, 12:12:32
 Job time : 11.7692 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:51:36 ; Search time 6.15385 Seconds
(without alignments)
169.228 Million cell updates/sec

Title: US-10-044-703-78

Perfect score: 96

Sequence: 1 IDELKTNSLTSILTYHV 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	96	100.0	1 MP70 MYCTU	Q50769 mycobacteri
2	76	79.2	2 MP83 MYCTU	Q10790 mycobacteri
3	46	47.9	1 Y831 SYN3	P74615 synchocyst
4	46	47.9	1 GPM1 HELPJ	Q92km7 helicobacte
5	45	46.9	1 P100 HSV6U	Q00701 human harpe
6	43	44.8	1 YH35 SYN3	P73392 synchocyst
7	43	44.8	1 ROLT_AGRH	P15397 agrobacteri
8	43	44.8	1 YFIB_BACSU	P24718 bacillus su
9	42	43.8	1 GLO2 CALJA	Q83333 callithrix
10	42	43.8	1 PERE_ARMRU	P59121 armoracia r
11	42	43.8	1 HEM2 SYN3	P77969 synchocyst
12	42	43.8	1 EM70 YEAST	P32802 saccharomyc
13	42	43.8	1 TMS6 MOUSE	Q9db10 mus musculu
14	42	43.8	1 VDP HUMAN	O60763 homo sapien
15	41.5	43.2	1 SPS_SOLTU	Q33845 solanum tub
16	41	42.7	1 YH35 SYN2	O33752 synchococc
17	41	42.7	1 PER3 ARMRU	P7180 armoracia r
18	40.5	42.2	1 LFCA SYN3	Q55797 synchocyst
19	40	41.7	1 EFID PHOLL	Q9v118 drosophila
20	40	41.7	1 EFTS PHOLL	O7n8p6 photorhabdu
21	40	41.7	1 PERC_ARMRU	P25233 armoracia r
22	40	41.7	1 PE22 ARATH	P24102 arabidopsis
23	40	41.7	1 V263 ASF87	P23164 african swi
24	40	41.7	1 F06C ARATH	P46312 arabidopsis
25	40	41.7	1 GAA1 CHICK	P19150 gallus gall
26	40	41.7	1 GAA1 MOUSE	P18504 mus musculu
27	40	41.7	1 GAA1 BOVIN	P08219 bos taurus
28	40	41.7	1 GAA1 HUMAN	P14867 homo sapien
29	39.5	41.1	1 PVRC_AGRF5	Q8ui99 agrobacteri
30	39.5	41.1	1 YC44 GUTH	O78437 guillardia
31	39	40.6	1 Y050 RICO	Q92j17 rickettsia
32	39	40.6	1 Y487 THEAC	Q9hkv6 thermoplasm
33	39	40.6	1 RPOA_MYCPN	Q50295 mycoplasma

34	39	40.6	340	1	CWFH SCHPO	O94620 schizosacch
35	39	40.6	373	1	RAPJ_BACSU	O34327 bacillus su
36	39	40.6	404	1	HUTI_THEVO	O978n3 thermoplasm
37	39	40.6	769	1	MLHI_YEAST	P38920 saccharomyc
38	39	40.6	1849	1	T172_HUMAN	O14991 homo sapien
39	38.5	40.1	373	1	Y812_METJA	Q58222 methanococc
40	38.5	40.1	380	1	Y079_METJA	O60389 methanococc
41	38.5	40.1	1054	1	SPS1_CRAPL	O04932 craterostig
42	38	39.6	203	1	Y818_METJA	Q58222 methanococc
43	38	39.6	206	1	CYSC_BUCAI	P57497 buchnera ap
44	38	39.6	276	1	Y415_CHLTR	O84420 chlamydia t
45	38	39.6	324	1	B3G2_RAT	Q92137 rattus norv

ALIGNMENTS

RESULT 1

MP70 MYCTU
ID MP70 MYCTU STANDARD; PRT; 193 AA.
AC Q50769; Q48934; Q48946; Q48947; Q48948; Q50656;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Immunogenic protein MPT70 precursor.
GN MPT70 OR MPB70 OR RV2875 OR MT3943 OR MTCY274.06 OR MB2900.
OS Mycobacterium tuberculosis, and
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=95176:87; PubMed=7871388;
RA Matsumoto S., Matsuo T., Ohara N., Hotokezaka H., Naitoh M.,
RA Minami J., Yamada T.;
RT "Cloning and sequencing of a unique antigen MPT70 from Mycobacterium
tuberculosis H37Rv and expression in BCG using E. coli-mycobacteria
shuttle vector.";
RL Scand. J. Immunol. 41:281-287 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Siglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Usterback T., Weidman J.C., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains".
RL J. Bacteriol. 184:5479-5490 (2002).
RN [4]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC SPECIES=M.bovis; STRAIN=BCG / Pasteur, and BCG / Tokyo;
RX MEDLINE=89306542; PubMed=2663636;
RA Terasaka K., Yamaguchi R., Matsuo K., Yamazaki A., Nagai S.,

RA Yamada T.;
 RT "Complete nucleotide sequence of immunogenic protein MPB70 from
 RL Mycobacterium bovis BCG";
 RN FEMS Microbiol. Lett. 49:273-276(1989).
 [5]
 RN SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=BCG / Pasteur, and BCG / Tokyo;
 RX MEDLINE=90218009; PubMed=1691265;
 RA Radford A., Wood P., Billman-Jacobe H., Geysen H., Mason T.,
 RA Tribick G.;
 RT "Epitope mapping of the Mycobacterium bovis secretory protein MPB70
 RT using overlapping peptide analysis";
 RL J. Gen. Microbiol. 136:265-272(1990).
 [6]
 RN SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=BCG / Pasteur, and BCG / Tokyo;
 RX MEDLINE=96004459; PubMed=7551028;
 RA Takemitsu M., Matsumoto S., Ohara N., Kitaura H., Mizuno A.,
 RA Yamada T.;
 RT "Differential transcription of the MPB70 genes in two major groups of
 RT Mycobacterium bovis BCG substrains";
 RL Microbiology 141:1601-1607(1995).
 [7]
 RN SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Mensepe C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 [8]
 RN SEQUENCE OF 49-193 FROM N.A.
 RC SPECIES=M.bovis;
 RX MEDLINE=88153076; PubMed=3278986;
 RA Radford A.J., Duffield B.J., Plackett P.;
 RT "Cloning of a species-specific antigen of Mycobacterium bovis";
 RL Infect. Immun. 56:921-925(1988).
 CC -!- SUBUNIT: GENERALLY FOUND AS A MONOMER; HOMODIMER IN CULTURE
 CC FLUIDS.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MISCELLANEOUS: PRODUCED IN HIGH CONCENTRATION BY BCG TOKYO,
 CC MOREAU, RUSSIA AND SWEDEN (HIGH-PRODUCER SUBSTRAINS), WHEREAS IN
 CC BCG PASTEUR, COPENHAGEN AND TICE (LOW-PRODUCER SUBSTRAINS) IT IS
 CC DETECTED AT 1% (W/W) OR LESS OF THE CONCENTRATION OF BCG TOKYO.
 CC THE DIFFERENCE IN THE SECRETION BETWEEN BCG TOKYO AND PASTEUR WAS
 CC ATTRIBUTED TO DIFFERENTIAL TRANSCRIPTION EFFICIENCIES.
 CC -!- SIMILARITY: Contains 1 FAS1 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D37968; BAA07184.1; -
 DR EMBL: Z74024; CAA98373.1; -
 DR EMBL: AE007118; AAK47268.1; -
 DR EMBL: D38229; BAA07402.1; -
 DR EMBL: M33916; AAA25366.1; -
 DR EMBL: D38230; BAA07403.1; -
 DR EMBL: EX248344; CAD96587.1; -
 DR EMBL: A08199; CAA00760.1; -
 DR EMBL: M37840; AAA25355.1; -
 DR PIR: A37195; A37195.
 DR PIR: A48320; A48320.
 DR PIR: F70923; F70923.
 DR TIGR: MT2943; -
 DR Tuberculist; RV2875; -
 DR InterPro; IPR000782; BIGH3_FAS1.

DR Pfam: PF02469; Fasciclin; 1.
 DR SMART: SMO0554; FAS1; 1.
 DR PROSITE: PS0213; FAS1; 1.
 KW Antigen; Signal; Complete proteome.
 FT SIGNAL 1 30
 FT CHAIN 31 193 IMMUNOGENIC PROTEIN MPT70.
 FT DOMAIN 57 189 FAS1.
 FT CONFLICT 16 17 GL -> AV (IN REF. 6).
 FT CONFLICT 101 101 P -> R (IN REF. 8).
 SQ SEQUENCE 193 AA; 19072 MW; 228695731C3FFB00 CRC64;
 Query Match 100.0%; Score 96; DB 1; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 IDELKTNSLLTSLTYHVV 20
 Db 114 IDELKTNSLLTSLTYHVV 133
 RESULT 2
 MP83 MYCTU STANDARD; PRT; 220 AA.
 ID MP83 MYCTU
 AC Q10730; P71493;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Cell surface lipoprotein MPT83 precursor (Lipoprotein P23).
 GN MPT83 OR MP83 OR RV2873 OR MT2940 OR MTCV274.04 OR MB2898.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1773, 1765;
 [1]
 RN SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=96233689; PubMed=8633206;
 RA Hewinson R.G., Mitchell S., Russell W.P., McAdam R.A.,
 RA Jacobs W.R. Jr.;
 RT "Molecular characterization of MPT83: a seroreactive antigen of
 RT Mycobacterium tuberculosis with homology to MPT70";
 RL Scand. J. Immunol. 43:490-499(1996).
 [2]
 RN SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornesby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence";
 RL Nature 393:537-544(1998).
 [3]
 RN SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL J. Bacteriol. 184:5479-5490(2002).
 [4]
 RN SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=BCG / Tokyo;

P74615;
30-MAY-2000 (Rel. 39, Created)
30-MAY-2000 (Rel. 39, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein sll1483 precursor.
GN SLL1483
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
NCBI_TaxID=1148;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
MYAJOHJI N., Hiroseawa M., Sugihara M., Sasamoto S., Kimura T.,
HOBUUCHI T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
Synechocystis sp. strain PC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
RC -!- SIMILARITY: Contains 1 FAS1 domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to licenses@lsb-sib.ch).

DR EMBL; D90916; BAU18723.1; -;
PIR; S76811; S76811.
DR InterPro; IPR000782; BiG3 FAS1.
DR Pfam; PF02469; Fasciclin_1.
DR SMART; SMC0554; FAS1; 1.
DR PROSITE; PS0213; FAS1; 1.
KW HYPOTHETICAL protein; Signal; Complete proteome.
FT SIGNAL 1 26 POTENTIAL,
FT CHAIN 27 180 HYPOTHETICAL PROTEIN SLL1483.
FT DOMAIN 45 176 FAS1
SQ SEQUENCE 180 AA; 18471 MW; 1DE2D9BAE8EE389 CRG64;

Query Match 47.9%; Score 46; DB 1; Length 180;
Best Local Similarity 58.8%; Pred.No. 2.2;
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 LKTNSSLSLTLYHYV 20
DB 103 LPENCKLVKITRYHV 119

RESULT 4
GPML_HELPJ
IC GPML_HELPJ STANDARD; PRT; 491 AA.

ID QZKH7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DE_3-B-phosphoglycerate-independent phosphoglycerate mutase
DE [ec 5.4.2.1] (Phosphoglyceromutase) (BPG-INDEPENDENT PGAM) (IPGM).
GN GPML OR PGM OR JHP0908
OS Halobacter pylori 399 (Campylobacter pylori 399).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
NCBI_TaxID=85963;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923582;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., Dejonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nichelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,

```

RX MEDLINE=94118404; PubMed=8289364;
RA Nicholas J., Martin M.;
RT "Nucleotide sequence analysis of a 38.5-kilobase-pair region of the
RT genome of human herpesvirus 6 encoding human cytomegalovirus
RT immediate-early gene homologs and transactivating functions.";
RL J. Virol. 68:597-610(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95266321; PubMed=7747482;
RA Gompels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
RT Martin M.B., Efsthathiou S., Craxton M., Macaulay H.A.;
RT "The DNA sequence of human herpesvirus-6: structure, coding content,
RT and genome evolution.";
RL Virology 209:129-51(1995).
CC -!- SIMILARITY: TO THE LARGE STRUCTURAL PHOSPHOPROTEINS OF HSV-7 AND
CC HCMV UL32.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M87287; AAA46012.1; -
CC EMBL; L25528; AAA16716.1; ALT_INIT.
CC EMBL; X83413; CAA58438.1; -
CC Matrix protein; Phosphorylation.
CC SEQUENCE 870 AA; 97071 MW; F25954DEA19BF824 CRC64;
CC -----
DR DQ Match 46.9%; Score 45; DB 1; Length 870;
DR Best Local Similarity 50.0%; Pred. No. 18;
DR Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
Ddb 1 IDELKTNSLLTSLTYHYV 20
:||||:|||||
54 VDDLKTLVNLVLWLHYV 73

RESULT 6
YH35_SVNY3
IID YH35_SVNY3 STANDARD; PRT; 133 AA.
AC P73392;
DDT 30-MAY-2000 (Rel. 39, Created)
DDT 30-MAY-2000 (Rel. 39, Last sequence update)
DDT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein sll1735.
GN SLL1735.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asanizu E., Nakamura Y.,
RT Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RT Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RT Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- SIMILARITY: Contains 1 FAS1 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).

```

```

CC      EMBL; D90906; BAA17432.1; --
DR      PIR; S77329; S77329. BIGH3_FAS1.
DR      InterPro; IPR000782; BIGH3_FAS1.
DR      Pfam; PF02469; Fasciclin_1.
DR      SMART; SM00554; FAS1; 1.
DR      PROSITE; PS0213; FAS1; 1.
KW      Hypothetical protein; Complete proteome.
FT      DOMAIN 1 130 FAS1.
SQ      SEQUENCE 133 AA; 14105 MW; 427DFPB427D85C4C CRC64;

Query Match      44.8%; Score 43; DB 1; Length 133;
Best Local Similarity 55.0%; Pred. No. 5.1;
Matches 11; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      1 IDELKTNSLLTSLITYHV 20
DB      54 IITLVQNIQLARILTYHV 73

RESULT 7
ROUT_AGRH      STANDARD;      PRT;      274 AA.
AC      P15397.
DT      01-APR-1990 (Rel. 14, Created)
DT      01-APR-1990 (Rel. 14, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      ROL B (TR) protein.
GN      ROLB.
OS      Agrobacterium rhizogenes.
OC      Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC      Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX      NCBI_TaxID=359;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A4;
RX      MEDLINE=91346650; PubMed=2102840;
RA      Bouchey D., Camilleri C.;
RT      "Identification of a putative rol B gene on the TR-DNA of the
RT      Agrobacterium rhizogenes A4 Ri plasmid.";
RL      Plant Mol. Biol. 14:617-619(1990).
CC      -!- FUNCTION: IT CONTRIBUTES TO THE ROOT INDUCING ACTIVITY.
CC      -!- SIMILARITY: TO THE ROL B PROTEIN, AND TO THE N-TERMINAL OF THE
CC      RI TL-DNA ORF 8 PROTEIN.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL; X15952; CAA34076.1; --
DR      EMBL; X15952; CAA34077.1; ALT_INIT.
DR      PIR; S11872; S11872.
DR      InterPro; IPR008992; Bact_endotox.
DR      InterPro; IPR006064; Glycosidase.
DR      Pfam; PF02027; ROLB_RolC; 1.
KW      Plasmid.
SQ      SEQUENCE 274 AA; 31253 MW; FD17652A83BF2F55 CRC64;

Query Match      44.8%; Score 43; DB 1; Length 274;
Best Local Similarity 38.9%; Pred. No. 11;
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY      1 IDELKTNSLLTSLITYH 18
DB      79 VNLQKINAEILTRKVCYH 96

RESULT 8

```

```

YFIB_BACSU
ID      YFIB_BACSU      STANDARD;      PRT;      573 AA.
AC      P54718;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Hypothetical ABC transporter ATP-binding protein yfIB.
GN      YFIB OR BSU08210.
DB      Bacillus subtilis.
OS      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX      NCBI_TaxID=1423;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=168;
RX      MEDLINE=96262713; PubMed=8704981;
RA      Yamamoto H., Uchiyama S., Fajar A.N., Ogasawara N., Sekiguchi J.;
RT      "Determination of a 12 KB nucleotide sequence around the 76 degrees
RT      region of the Bacillus subtilis chromosome.";
RL      Microbiology 142:1417-1421(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=168;
RX      MEDLINE=98044033; PubMed=9384377;
RA      Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA      Azevedo V., Bertsch M.G., Bessieres P., Bolotin A., Borcherst S.,
RA      Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA      Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA      Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA      Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA      Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA      Fitts C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA      Ghim S.Y., Glaser P., Goiffau A., Gollightly E.J., Grandi G.,
RA      Guiseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA      Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA      Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA      Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA      Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA      Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA      Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA      Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA      Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA      Presecan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
RA      Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA      Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA      Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA      Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA      Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA      Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA      Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA      Winters P., Wipat A., Yamamoto H., Yanane K., Yasumoto K., Yata K.,
RA      Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT      "The complete genome sequence of the Gram-positive bacterium Bacillus
RT      subtilis.";
RL      Nature 390:249-256(1997).
CC      -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC      -!- SIMILARITY: Belongs to the ABC transporter family. MSBA subfamily.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL; D50543; BAA09106.1; --
DR      EMBL; Z99108; CAB2650.1; --
DR      PIR; E69802; E69802.
DR      HSP; P13569; INBD.
DR      Subtilist; BG11849; yfIB.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR001140; ABC_TM_transp.
DR      InterPro; IPR003439; ABC_transporter.

```

DR Pfam; PF00664; ABC membrane; 1.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS0929; ABC TM1F; 1.
 DR PROSITE; PS0211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS0893; ABC_TRANSPORTER_2; 1.
 DR Hypothetical protein; ATP-binding; Transport; Transmembrane;
 KW Complete proteome.
 FT TRANSMEM 17 37 POTENTIAL.
 FT TRANSMEM 52 72 POTENTIAL.
 FT TRANSMEM 127 147 POTENTIAL.
 FT TRANSMEM 153 173 POTENTIAL.
 FT TRANSMEM 238 258 POTENTIAL.
 FT TRANSMEM 275 295 POTENTIAL.
 FT NP_BIND 364 371 ATP (POTENTIAL).
 SQ SEQUENCE 573 AA; 63941 MW; E2BIAA4557E4B71 CRC64;

Query Match 44.8%; Score 43; DB 1; Length 573;
 Best Local Similarity 50.0%; Pred. No. 25;
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 ELKTNSSLLTSILTVH 18

Db 499 DLQTEAKLEAISTVH 514

RESULT 9

GL02_CALJA STANDARD; PRT; 260 AA.
 ID_GL02_CALJA
 AC Q28333;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hydroxyl-glutathione hydrolase (EC 3.1.2.6) (Glyoxalase II)
 DE (Glx II) (Germ cell specific protein).
 GN HGH.
 OS Callitrix jacchus (Common marmoset).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;
 OC Callitrich.
 OX NCBI_TaxID=9483;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Saunders P.T.K., Gaughan J.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Thioesterase that catalyzes the hydrolysis of S-D-
 CC lactoyl-glutathione to form glutathione and D-lactic acid.
 CC -!- CATALYTIC ACTIVITY: (S)-(2-hydroxyacetyl)glutathione + H(2)O =
 CC glutathione + a 2-hydroxy acid anion.
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
 CC -!- PATHWAY: glyoxal pathway.
 CC -!- TISSUE SPECIFICITY: Testis.
 CC -!- SIMILARITY: Belongs to the glyoxalase II family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X95294; CAA64612.1; -.
 DR HSP; O16775; LQH5.
 DR InterPro; IPR001279; Blactmase-like.
 DR Pfam; PF00753; lactamase_B; 1.
 KW Hydrolase; Zinc.
 FT METAL 54 54 ZINC 1 (BY SIMILARITY).
 FT METAL 56 56 ZINC 1 (BY SIMILARITY).
 FT METAL 58 58 ZINC 2 (BY SIMILARITY).
 FT METAL 59 59 ZINC 2 (BY SIMILARITY).

FT METAL 110 110 ZINC 1 (BY SIMILARITY).
 FT METAL 134 134 ZINC 1 AND 2 (BY SIMILARITY).
 FT METAL 173 173 ZINC 2 (BY SIMILARITY).
 SQ SEQUENCE 260 AA; 28792 MW; 11BB1D4C4B1AF962A CRC64;
 Query Match 43.8%; Score 42; DB 1; Length 260;
 Best Local Similarity 38.9%; Pred. No. 16;
 Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 IDELKNSSLLTSILTVH 18

Db 37 VERAKKHGVMLTTLTVH 54

RESULT 10

PERE_ARMRU STANDARD; PRT; 306 AA.
 ID_PERE_ARMRU
 AC P59131;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Peroxidase E5 (EC 1.11.1.7).
 GN HRPES.
 OS Amoracia rusticana (Horse radish) (Amoracia laphatifolia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Armoracia.
 OX NCBI_TaxID=3704;
 RN [1]
 RP SEQUENCE, AND X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 RA Morita Y., Mikami B., Yamashita H., Lee J.Y., Aibara S., Sato M.,
 RA Katsube Y., Tanaka N.;
 RT "Primary and crystal structures of horseradish peroxidase isozyme
 RT E5.";
 RL (In) Lobarzewski J., Greppin H., Penel C., Gaspar T. (eds.);
 RL Biochemical, molecular, and physiological aspects of plant
 RL peroxidases, pp.81-88, University of M. Curie-Skłodowska and
 RL University of Geneva, Lublin and Geneva (1991).
 CC -!- FUNCTION: Removal of H(2)O(2), oxidation of toxic reductants,
 CC biosynthesis and degradation of lignin, suberization, auxin
 CC catabolism, response to environmental stresses such as wounding,
 CC pathogen attack and oxidative stress. These functions might be
 CC dependent on each isozyme/isoform in each plant tissue.
 CC -!- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
 CC -!- COFACTOR: Binds 1 heme B (iron-protoporphyrin IX) group and 2
 CC calcium ions per subunit.
 CC -!- SIMILARITY: Belongs to the peroxidase family. Classical plant
 CC (class III) peroxidase subfamily.
 DR HSP; P00433; 2ATU.
 DR InterPro; IPR002016; Peroxidase.
 DR Pfam; PF00141; peroxidase; 1.
 DR PRINTS; PR00458; PEROXIDASE.
 DR PROSITE; PS00435; PEROXIDASE_1; 1.
 DR PROSITE; PS00436; PEROXIDASE_2; 1.
 DR PROSITE; PS00873; PEROXIDASE_4; 1.
 KW Oxidoreductase; Glycoprotein; Peroxidase; Iron; Heme;
 KW Multigene family; Calcium; Pyridoxine carboxylic acid.
 FT METAL 43 43 CALCIUM 1 (BY SIMILARITY).
 FT METAL 45 46 CALCIUM 1 (VIA CARBONYL OXYGEN)
 FT METAL 48 48 CALCIUM 1 (BY SIMILARITY).
 FT METAL 50 50 CALCIUM 1 (VIA CARBONYL OXYGEN)
 FT METAL 52 52 CALCIUM 1 (BY SIMILARITY).
 FT METAL 171 171 CALCIUM 1 (BY SIMILARITY).
 FT METAL 222 222 CALCIUM 2 (BY SIMILARITY).
 FT METAL 225 225 CALCIUM 2 (BY SIMILARITY).
 FT METAL 230 230 CALCIUM 2 (BY SIMILARITY).
 FT ACT_SITE 38 38 BY SIMILARITY.
 FT ACT_SITE 42 42 DISTAL HISTIDINE.
 FT ACT_SITE 70 70 HYDROGEN-BOND (BY SIMILARITY).
 FT ACT_SITE 139 139 SUBSTRATE BINDING (BY SIMILARITY).
 FT METAL 170 170 IRON (HEME AXIAL LIGAND).

FT ACT SITE 247 247 HYDROGEN-BOUND (BY SIMILARITY).
FT DISULFID 11 91 BY SIMILARITY.
FT DISULFID 44 49 BY SIMILARITY.
FT DISULFID 97 300 BY SIMILARITY.
FT DISULFID 177 209 BY SIMILARITY.
FT MOD RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 306 AA; 33722 MW; ABA344197078FE6 CRC64;

Query Match 43.8%; Score 42; DB 1; Length 306;
Best Local Similarity 44.4%; Pred. No. 19;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYH 18
DB 23 VDELOTDPRIASILRLH 40
:|||:|:|:|

RESULT 11
HEM2_SVNY3
ID_HEM2_SVNY3 STANDARD; PRT; 327 AA.
AC E77969;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Delta-aminolevulinic acid dehydratase (EC 4.2.1.24) (Porphobilinogen synthase) (ALAD) (ALADH).
GN HEMB OR SL11994.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX KANeko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y., Miyajima N., Hirosewa M., Sugliura M., Sasamoto S., Kimura T., Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S., Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yaeuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- CATALYTIC ACTIVITY: 2 5-aminolevulinate = porphobilinogen + 2 H(2O).
CC -!- COFACTOR: Zinc (By similarity).
CC -!- PATHWAY: Porphyrin biosynthesis; second step.
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SIMILARITY: Belongs to the ALADH family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; D90311; BAA18067.1;
CC PIR; S75506; S75506.
CC HSSP; P15002; 1B4E.
CC InterPro; IPR001731; ALAD_dehydratase.
CC Pfam; PF00490; ALAD; 1
CC PRINTS; PR00444; DALDHYDRATASE.
CC ProDom; PD002304; Alad_dehydratase; 1.
CC PROSITE; PS00169; D_ALA_DEHYDRATASE; 1.
CC Porphyrin biosynthesis; Heme biosynthesis; Lyase; Zinc;
KW Complete proteome.
FT DOMAIN 116 134 ZINC-BINDING (BY SIMILARITY).
FT ACT SITE 251 251 ZINC-BINDING.
SQ SEQUENCE 327 AA; 36163 MW; A7FAA7D5915300C9 CRC64;

Query Match 43.8%; Score 42; DB 1; Length 327;
Best Local Similarity 46.2%; Pred. No. 20;
Matches 12; Conservative 0; Mismatches 6; Indels 8; Gaps 1;

QY 1 IDELKTNSLLTS-----ILTYH 18
DB 292 IDEQVLTLETLSFKRAGADLIITYH 317
:|||:|:|:|

RESULT 12
EM70_YEAST
ID_EM70_YEAST STANDARD; PRT; 667 AA.
AC P32802; Q12101;
DT 01-OCT-1993 (Rel. 27, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endosomal P24A protein precursor (70 kDa endomembrane protein) (Phenome alpha-factor transporter) (Acidic 24 kDa late endocytic intermediate component).
DE EMP70 OR YLR083C OR I9449.11.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX Singer-Krueger B., Krueger U., Riesen H.;
RL Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansoorge W., Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A., Entian K.-D., Floeth M., Goifeau A., Hebling U., Heumann K., Hauss-Neitzel D., Hilbert H., Hilger F., Kleine K., Kleiner P., Louis E.J., Messenguy F., Mewes H.-W., Micsoga T., Moestl D., Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M., Portetelle D., Purnelle B., Reckmann S., Rieger M., Rinke M., Rose M., Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S., Underwood A.P., Urrestarazu L.A., Vandenberg M., Verhasselt P., Vierendeels F., Voet M., Voickaert G., Voss H., Wambutt R., Wedler E., Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL Nature 387:87-90(1997).
RN [3]
RP SEQUENCE OF 23-42.
RX STRAIN=RH732;
MEDLINE=93300835; PubMed=8314797;
RA Singer-Krueger B., Frank R., Crausaz F., Riezman H.;
RT "Partial purification and characterization of early and late endosomes from yeast. Identification of four novel proteins.";
RL J. Biol. Chem. 268:14376-14386(1993).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endosome (Probable).
CC -!- SIMILARITY: Belongs to the nonaspanin (TM9SF) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; X67316; CAA47730.1;
CC EMBL; U53880; AAB67587.1;
CC EMBL; Z73255; CAA97643.1;
CC GenOnline; 142145;
CC SGD; S0004073; EMP70.
CC InterPro; IPR004240; EMP70.
CC Pfam; PF02990; EMP70; 1.
KW Signal; Transmembrane; Glycoprotein.

```

FT SIGNAL 1 22
FT CHAIN 23 250 ENDOSOMAL P24A PROTEIN.
FT PROPP 251 667 POTENTIAL.
FT TRANSMEM 303 323 POTENTIAL.
FT TRANSMEM 371 391 POTENTIAL.
FT TRANSMEM 406 426 POTENTIAL.
FT TRANSMEM 443 463 POTENTIAL.
FT TRANSMEM 467 487 POTENTIAL.
FT TRANSMEM 528 548 POTENTIAL.
FT TRANSMEM 561 581 POTENTIAL.
FT TRANSMEM 597 617 POTENTIAL.
FT TRANSMEM 637 657 POTENTIAL.
FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 207 207 R -> C (IN REF. 1).
FT CONFLICT 268 268 G -> S (IN REF. 1).
FT CONFLICT 488 488 L -> S (IN REF. 1).
SQ SEQUENCE 667 AA; 75962 MW; B5B8A6876C548CA CRC64;

Query Match 43.8%; Score 42; DB 1; Length 667;
Best Local Similarity 53.3%; Pred. No. 43;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 LKTNSSLLTSILTYH 18
Db 570 LITSSLVILLITYH 584

RESULT 13
TMS6_MOUSE STANDARD; PRT; 811 AA.
AC Q9DB10;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transmembrane protease, serine 6 (EC 3.4.21.-) (Matritpase-2).
GN TMPRSS6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=C57BL/6J;
RX MEDLINE=22755759; PubMed=12744720;
RA Hooper J.D., Campagnolo L., Goodarzi G., Truong T.N., Stuhlmann H.,
RA Quigley J.P.;
RT "Mouse matritpase-2: identification, characterization and comparative
RT mRNA expression analysis with mouse hepsin in adult and embryonic
RT tissues.";
RL Biochem. J. 373:689-702 (2003).
[2]

SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotohori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schraml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmel S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Ienhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., Mckenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sadelin A., Schneider C., Semple C.A., Setou M., Shinada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wanhlestedt C., Wang Y., Watanabe Y., Wells C.,

```

```

RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
[3]
SEQUENCE FROM N.A.
RP TISSUE=Liver;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshlyuk S., Carninci P., Prange C.,
RA Raha S.S., Iaquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[4]
REVIEW.
RX MEDLINE=22668120; PubMed=12784999;
RA Netzel-Arnett S., Hooper J.D., Szabo R., Madison E.L., Quigley J.P.,
RA Bugge T.H., Antalis T.M.;
RT "Membrane anchored serine proteases: a rapidly expanding group of cell
RT surface proteolytic enzymes with potential roles in cancer.";
RL Cancer Metastasis Rev. 22:237-258 (2003).
CC -!- FUNCTION: May play a specialized role in matrix remodeling
CC processes in liver (By similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed at highest levels in liver, kidney
CC and uterus.
CC -!- DEVELOPMENTAL STAGE: Expressed at higher levels from 12.5 dpc to
CC 15.5 dpc with a peak at 13.5 dpc. Expression in the developing
CC liver as well as a restricted set of embryonic epithelial cells of
CC the nasal cavity and pharyngo-tympanic tubes.
CC -!- SIMILARITY: Belongs to peptidase family S1.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 3 LDL-receptor class A domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; AY240929; AAP69827.1; -
EMBL; AK004939; BAB23684.2; -
EMBL; BC029645; AAH29645.2; -
HSSP; P00763; 1DPO.
MEROPS; S01.308; -
MGD; MGI:1919003; TMPRSS6.
InterPro; IPR000859; CUB.
InterPro; IPR009003; Cys_ser_trypsin.
InterPro; IPR002172; LDL_receptor_A.

```



```

Query Match      43.8%; Score 42; DB 1; Length 962;
Best Local Similarity 62.5%; Pred. No. 64;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      1 IDELKTNSILTSILT 16
      |:|:|:|:|:|:|:|:|:|
DDB      739 IEELKRNQELLOSQT 754

RESULT 15
SPS_SOLTU
IID_SPS_SOLTU      STANDARD;      PRT; 1053 AA.
AC      Q43845;
DDT      15-DEC-1998 (Rel. 37, Created)
DDT      15-DEC-1998 (Rel. 37, Last sequence update)
DDT      15-DEC-1998 (Rel. 37, Last annotation update)
DE      Sucrose-phosphate synthase (EC 2.4.1.14) (UDP-glucose-fructose-
phosphate glucosyltransferase).
GN      Solanum tuberosum (Potato).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC      lamnids; Solanales; Solanaceae; Solanum.
OC      NCBI_TaxID=4113;
OC      [1]
PRP      SEQUENCE FROM N.A.
RC      STRAIN=cv. Desiree; TISSUE=Leaf;
RC      MEDLINE=35201832; PubMed=7894514;
RC      Zrenner R., Sallanoubat M., Willmitzer L., Sonnnewald U.;
RC      "Evidence of the crucial role of sucrose synthase for sink strength
RC      using transgenic potato plants (Solanum tuberosum L.).";
RRL      Plant J. 7:97-107(1995)
CCC      -1- FUNCTION: Involved in the regulation of carbon partitioning in the
CCC      leaves of plants. May regulate the synthesis of sucrose and
CCC      therefore play a major role as a limiting factor in the export of
CCC      photoassimilates out of the leaf.

```

```

CC -|- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
CC sucrose 6-phosphate.
CC -|- ENZYME REGULATION: Activity regulated by phosphorylation and
CC moderated by concentration of metabolites and light.
CC -|- PATHWAY: Sucrose synthesis.
CC -|- SUBUNIT: Homodimer or homotetramer (By similarity).
CC -|- PTM: Phosphorylated. However, phosphorylation is not essential for
CC enzyme function (By similarity).
CC -|- SIMILARITY: Belongs to the glycosyltransferase family 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73477; CAA51872.1; -
DR PIR; S34172; S34172.
DR InterPro; IPR001296; Glyco_transf_1.
DR Pfam; PF00534; Glycos_transf_1; 1.
KW Transferase; Glycosyltransferase; Phosphorylation.
SQ SEQUENCE 1053 AA; 118292 MW; D6C933798567B20A CRC64;

```

```

Query Match . 43.2%; Score 41.5; DB 1; Length 1053;
Best Local Similarity 68.8%; Pred. No. 86;
Matches 11; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

```

```

QY 1 IDEL-KTNSSLTSL 15
DB 517 IDEMSSTNSALLSL 532

```

```

Search completed: March 10, 2004, 12:06:35
Job time : 7.15385 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:57:36 ; Search time 31.6667 Seconds
(without alignments)
199.275 Million cell updates/sec

Title: US-10-044-703-78

Perfect score: 96

Sequence: 1 IDELKTNSLLTSILTYHV 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	96	100.0	226	Q9RMU8	Q9RMU8 mycobacteri
2	76	79.2	106	Q9RMU9	Q9RMU9 mycobacteri
3	57	59.4	219	Q9RD45	Q9RD45 streptomyc
4	55	57.3	3255	Q9PKM8	Q9PKM8 chlamydia m
5	53	55.2	719	Q8PTS4	Q8PTS4 methanosarc
6	53	55.2	877	Q8TQ29	Q8TQ29 methanosarc
7	52	54.2	151	Q8PSH3	Q8PSH3 methanosarc
8	51	53.1	167	Q877K8	Q877K8 bradyrhizob
9	51	53.1	184	Q89X15	Q89X15 bradyrhizob
10	51	53.1	311	Q9NH96	Q9NH96 anthopleura
11	51	53.1	343	Q9VAG8	Q9VAG8 anthocidari
12	51	53.1	344	Q9VAG9	Q9VAG9 anthocidari
13	49	51.0	133	Q8YX95	Q8YX95 anabaena sp
14	49	51.0	160	Q9R9N9	Q9R9N9 rhizobium m
15	49	51.0	160	Q92ZA8	Q92ZA8 rhizobium m
16	48	50.0	215	Q82C20	Q82C20 streptomyc

17	48	50.0	328	12	Q914J0
18	47.5	49.5	1139	4	Q9P278
19	47	49.0	159	5	Q86LW3
20	47	49.0	263	17	Q8TPB7
21	47	49.0	623	16	Q9RXB6
22	46.5	48.4	499	17	Q8TU20
23	46.5	48.4	506	17	Q8PWP5
24	46.5	48.4	825	11	Q80TD3
25	46	47.9	266	16	Q8CS42
26	45	46.9	221	16	Q8FUF7
27	45	46.9	239	16	Q53836
28	45	46.9	241	2	Q9KW25
29	45	46.9	261	16	Q8YQM3
30	45	46.9	263	16	Q99TB5
31	45	46.9	263	16	Q8NW25
32	45	46.9	683	11	P82198
33	44.5	46.4	150	16	Q8P4P4
34	44	45.8	141	16	Q8YVN9
35	44	45.8	144	16	Q8RGE8
36	44	45.8	220	16	Q86667
37	44	45.8	1505	16	Q8KFW2
38	43.5	45.3	185	16	Q8PGB4
39	43.5	45.3	894	10	Q8VYW8
40	43.5	45.3	1047	10	Q8FY54
41	43	44.8	214	16	Q814W4
42	43	44.8	255	16	Q92G35
43	43	44.8	365	17	Q8YK45
44	43	44.8	722	10	Q7X997
45	43	44.8	775	5	Q8N0B2

ALIGNMENTS

RESULT 1

Q9RMU8 PRELIMINARY; PRT; 236 AA.
ID Q9RMU8
AC Q9RMU8;
CT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DE MPT70.
GN MPT70.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_taxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RA Juarez M.D., Torres A., Bigi F., Espitia C.;
RT "Mycobacterium tuberculosis mpt83 and dipZ/thioredoxin genes are part
of the same translational unit."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF189006; AAF13402.1;
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR000782; SIGH3_FAS1.
DR Pfam; PF02469; Fasciclin; I.
DR SMART; SM00554; FAS1; 1.
DR PROSITE; PS0213; FAS1; 1.
SQ SEQUENCE 236 AA; 22467 MW; 7AB8517F63A12751 CRC64;

Query Match 100.0%; Score 96; DB 2; Length 226;
Best Local Similarity 100.0%; Pred. No. 8.3e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IDELKTNSLLTSILTYHV 20

Db 147 IDELKTNSLLTSILTYHV 166

RESULT 2

Q9RMU9

Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H., Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M., Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S., Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S., Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S., Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K., Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J., Hopwood D.A.,

"Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)."; EMBL; AL939106; CAB62767.1; GO: GO:0007155; P:cell adhesion; IEA. InterPro; IPR000782; HIGH_FAS1. InterPro; IPR000437; Prok_lipoprot_S. Pfam; PF02459; Fasciclin; 1. SMART; SMO0554; FAS1; 1. PROSITE; PS00213; FAS1; 1. PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1. Lipoprotein; Complete proteome.

SEQUENCE 219 AA; 22255 MW; 863F97B1D0E4AF91 CRC64;

Query Match 59.4%; Score 57; DB 16; Length 219;
Best Local Similarity 50.0%; Pred.No. 0.27;
Matches 10; Conservative 6; Mismatches 4; Indels 0; Gaps

Cy 1 IDELKTNSLLTSILTYHVV 20
::: : ::::|
Db 138 LDKVLNDKDLNLTLYHVV 157

RESULT 4

G9PFX08 PRELIMINARY; PRT; 3255 AA.

ID AC Q9PKM8
AD AC Q9PKM8
DT DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Adherence factor.
GN TC0437.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
CX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPn / Ni99;
RX MEDLINE=2015235; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
PT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae Ax35.";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AE002311; AAP39291.1; -
DR PIR; G81702; G81702.
DR TIGR; TC0437; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001395; Aldo/ket_red.
DR InterPro; IPR007577; Gly_transf_sug.
DR InterPro; IPR006473; Peptidase_C58_yt.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF04488; Gly_transf_sug; 1.
DR TIGRFAMS; TIGR01586; yopr_cys_prot; 1.
DR PROSITE; PS00083; ALDO_KETO_REDUCTASE_3; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
SQ Complete proteome.
SQ SEQUENCE 3255 AA; 3681141 MW; 8F648038634AF23F CRC64;

Query Match 57.3%; Score 55; DB 16; Length 3255;
 Best Local Similarity 50.0%; Pred. No. 9.6;
 Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHV 20
 DB 1711 LDEMKGVDVLTFTKVL 1730

RESULT 5
 Q8PTS4
 ID Q8PTS4 PRELIMINARY; PRT; 719 AA.
 AC Q8PTS4
 DT 01-OCT-2002 (TREMELrel. 22, Created)
 DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Hypothetical protein MM2638.
 GN MM2638.
 OS Methanosarcina mazei (Methanosarcina frisia).
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2209;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Goel / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;
 RX MEDLINE=22120827; PubMed=12125824;
 RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
 RA Martinez-Arias R., Henne A., Wlezer A., Baeumer S., Jacobi C.,
 RA Brueggemann H., Lienard T., Christmann A., Boencke M., Steckel S.,
 RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
 RA Fritz H.-J., Gottschalk G.;
 RT "The genome of Methanosarcina mazei: evidence for lateral gene
 RT transfer between Bacteria and Archaea."
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
 DR EMBL; AE013509; AAM32334.1; -;
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR000782; BIGH3_FAS1.
 DR Pfam; PF02469; Fasciclin; 2.
 DR PROSITE; PS0213; FAS1; 2.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 719 AA; 77891 MW; 129E3141FE0003A1 CRC64;

Query Match 55.2%; Score 53; DB 17; Length 719;
 Best Local Similarity 47.4%; Pred. No. 4.3;
 Matches 9; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 2 DELKTNSSLLTSILTYHV 20
 DB 417 NQLMNTTLNQVLSYHV 435

RESULT 6
 Q8TQ29
 ID Q8TQ29 PRELIMINARY; PRT; 877 AA.
 AC Q8TQ29
 DT 01-JUN-2002 (TREMELrel. 21, Created)
 DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Hypothetical protein MAL723.
 GN MAL723.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
 RA Allen N., Navlor J., Stange-Thomann N., Dearellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Guss A.M.,

RA Hederich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umavam L.A., White O., White R.H., de Macario E.C.,
 RA Perry J.G., Jarell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity."
 RL Genome Res. 12:532-542(2002).
 DR EMBL; AE010846; AAM05130.1; -;
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR000782; BIGH3_FAS1.
 DR Pfam; PF02469; Fasciclin; 2.
 DR SMART; SM00554; FAS1; 2.
 DR PROSITE; PS0213; FAS1; 2.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 877 AA; 94651 MW; D8AF942EC0714AB4 CRC64;

Query Match 55.2%; Score 53; DB 17; Length 877;
 Best Local Similarity 50.0%; Pred. No. 5.3;
 Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 DELKTNSSLLTSILTYHV 19
 DB 585 DQLMNTTLRLKVLHYV 602

RESULT 7
 Q8PSH3
 ID Q8PSH3 PRELIMINARY; PRT; 151 AA.
 AC Q8PSH3
 DT 01-OCT-2002 (TREMELrel. 22, Created)
 DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Hypothetical protein MM3106.
 GN MM3106.
 OS Methanosarcina mazei (Methanosarcina frisia).
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2209;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Goel / G01 / ATCC BAA-199 / DSM 3647 / OCM 88;
 RX MEDLINE=22120827; PubMed=12125824;
 RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
 RA Martinez-Arias R., Henne A., Wlezer A., Baeumer S., Jacobi C.,
 RA Brueggemann H., Lienard T., Christmann A., Boencke M., Steckel S.,
 RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
 RA Fritz H.-J., Gottschalk G.;
 RT "The genome of Methanosarcina mazei: evidence for lateral gene
 RT transfer between Bacteria and Archaea."
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
 DR EMBL; AE013568; AAM32802.1; -;
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR000782; BIGH3_FAS1.
 DR Pfam; PF02469; Fasciclin; 1.
 DR PROSITE; PS0213; FAS1; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 151 AA; 15728 MW; 4DF72F819E90048F CRC64;

Query Match 54.2%; Score 52; DB 17; Length 151;
 Best Local Similarity 45.0%; Pred. No. 1.2;
 Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 IDELKTNSLLTSILTYHV 20
 DB 69 LEDLADLQALDVLTYHV 88

RESULT 8
 Q877K8
 ID Q877K8 PRELIMINARY; PRT; 167 AA.
 AC Q877K8;

```

DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE B1r2474 protein (B1r5191 protein).
GN B1r2474 OR B1r5191.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484938; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiuni T.,
RA Sasanoto S., Watanabe A., Iidesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Teurouka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AF005943; BAC47739.1; -.
DR EMBL; AF005954; BAC50456.1; -.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR000782; BIGH3_FAS1.
DR Pfam; PF02469; Fasciclin; I.
DR PROSITE; PS0213; FAS1; 1.
KW Complete proteome.
SQ SEQUENCE 167 AA; 17071 MW; D5330BE574564727 CRC64;

Query Match 53.1%; Score 51; DB 16; Length 167;
Best Local Similarity 71.4%; Pred. No. 2;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 7 NSSLLTSILTYHV 20
Db 91 NKAULTILTYHV 104

RESULT 9
Q89X15 PRELIMINARY; PRT; 184 AA.
AC Q89X15;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE B1l0507 protein.
GN B1l0507.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484938; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiuni T.,
RA Sasanoto S., Watanabe A., Iidesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Teurouka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AF005936; BAC45772.1; -.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR000782; BIGH3_FAS1.
DR Pfam; PF02469; Fasciclin; I.
DR SMART; SM00554; FAS1; 1.
DR PROSITE; PS0213; FAS1; 1.
KW Complete proteome.
SQ SEQUENCE 184 AA; 19104 MW; 628236FC60E8D6BB CRC64;

Query Match 53.1%; Score 51; DB 16; Length 184;
Best Local Similarity 71.4%; Pred. No. 2.2;

Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 7 NSSLLTSILTYHV 20
Db 91 NKAULTILTYHV 104

Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 7 NSSLLTSILTYHV 20
Db 104 NKATLTILTYHV 117

RESULT 10
Q9NH96 PRELIMINARY; PRT; 311 AA.
AC Q9NH96;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative cell adhesion protein Sym32.
GN Sym32.
OS Anthopleura elegantissima (Sea anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynanthese; Actiniidae; Anthopleura.
OX NCBI_TaxID=6110;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20371129; PubMed=10908850;
RA Reynolds W.S., Schwarz J.A., Weis V.M.;
RT "Symbiosis-enhanced gene expression in cnidarian-algal associations:
RT cloning and characterization of a cDNA, sym32, encoding a possible
RT cell adhesion protein."
RL Comp. Biochem. Physiol. 126:33-44(2000).
DR EMBL; AF229054; AAF65308.1; -.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR000782; BIGH3_FAS1.
DR Pfam; PF02469; Fasciclin; 2.
DR SMART; SM00554; FAS1; 2.
DR PROSITE; PS0213; FAS1; 2.
SQ SEQUENCE 311 AA; 32956 MW; E27CC6C1D4A1ACA9 CRC64;

Query Match 53.1%; Score 51; DB 5; Length 311;
Best Local Similarity 55.0%; Pred. No. 3.9;
Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 IDELKTNSLLTSILTYHV 20
Db 227 LKSLKNIPLLTKLYHV 246

RESULT 11
Q9UAG8 PRELIMINARY; PRT; 343 AA.
AC Q9UAG8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE EBP-beta precursor.
GN EBP-BETA.
OS Anthocidaris crassispina (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Echinoidae; Echinacea; Echinoida; Echinometridae;
OC Anthocidaris.
OX NCBI_TaxID=7629;
RN [1]
RP SEQUENCE FROM N.A.
RA Hirate Y., Tomita K., Yamamoto S., Kobari K., Uemura I., Yamasu K.,
RA Suyemitsu T.;
RT "Association of the sea urchin EGF-related peptide, EGIP-D, with
RT Fasciclin I-related ECM proteins from the sea urchin Anthocidaris
RT crassispina."
RL Dev. Growth Differ. 0:0-0(1999).
DR EMBL; AB024732; BAA82957.1; -.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR000782; BIGH3_FAS1.
DR Pfam; PF02469; Fasciclin; 2.
DR SMART; SM00554; FAS1; 2.
DR PROSITE; PS0213; FAS1; 2.

```

Yasuda M., Tabata S.;
"Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.",
DNA Res. 8:205-213(2001).
ENML; AP003585; BAB73277.1; --
PR; AE1971; AE1971.
GO; GO:0007155; P:cell adhesion; IEA.
InterPro; IPR000782; BIGH3_FAS1.
PFam; PF02469; Fasciclin; 1.
SMART; SM00554; FAS1; 1.
PROSITE; PS0213; FAS1; 1.
Hypothetical protein; Complete proteome.
SEQUENCE 133 AA; 13997 MW; ICEEDA2740CA7B4E CRC64;

Query Match 51.0%; Score 49; DB 16; Length 133;
Best Local Similarity 60.0%; Pred. No. 3.5;
Matches 12; Conservative 0; Mismatches 8; Indels 0; Gaps

QY 1 IDELNTSSLLTSILTYHV 20
DB 54 IQTLVQNIPQLTRILTHVV 73

RESULT 14

Q9R9N9 PRELIMINARY; PRT; 160 AA.

AC Q9R9N9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SM12C4.2.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rickettsiales; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=20511116; PubMed=11097914;
RA Davey M.E., de Bruijn F.J.; --
RT "A Homologue of the Tryptophan-Rich Sensory Protein TspO and FixL
Regulate a Novel Nutrient Deprivation-Induced Sinorhizobium meliloti
Locust";
RL Appl. Environ. Microbiol. 66:5353-5359 (2000).
DR ENML; AP179401; AAP01193.1; --
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR000782; BIGH3_FAS1.
DR Pfam; PF02469; Fasciclin; 1.
DR SMART; SM00554; FAS1; 1.
DR PROSITE; PS0213; FAS1; 1.
KW Hypothetical protein.
SQ SEQUENCE 160 AA; 16506 MW; 3F698ABDFF1304B2 CRC64;

Query Match 51.0%; Score 49; DB 2; Length 160;
Best Local Similarity 71.4%; Pred. No. 4.2;
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps

QY 7 NSSLLTSILTYHV 20
DB 84 NKQKLTILTYHV 97

RESULT 15

Q92ZA8 PRELIMINARY; PRT; 160 AA.

AC Q92ZA8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Nex18 Symbiotically induced conserved protein.
GN NEX18 OR RA0582 OR SMA1077.
OS Rhizobium meliloti (Sinorhizobium meliloti).

OG Plasmid pSymA (megaplasmid 1).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396509; PubMed=11481432;
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
 RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
 RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
 RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
 RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
 RT "Nucleotide sequence and predicted functions of the entire
 RL Sinorhizobium meliloti pSymA megaplasmid";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
 DR EMBL; AS007248; AAK5240.1; -.
 DR PIR; F95334; F95334.
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR000782; BIGH3_FAS1.
 DR Pfam; PF02469; Fasciclin; I.
 DR SMART; SM00554; FAS1; 1.
 DR PROSITE; PS0213; FAS1; 1.
 KW Plasmid; Complete proteome.
 SQ SEQUENCE 160 AA; 16482 MW; A8F107A3DFACAEED CRC64;

Query Match 51.0%; Score 49; DB 16; Length 160;
 Best Local Similarity 71.4%; Pred. No. 4.2;
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 7 NSSLTSTLTTHVV 20
 DB 84 NKQKLTETLTTHVV 97

Search completed: March 10, 2004, 12:11:01
 Job time : 33.7917 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:50:51 ; Search time 45.1859 Seconds
(without alignments)
118.807 Million cell updates/sec

Title: US-10-044-703-80

Perfect score: 95

Sequence: 1 GVSTANATVYMIDSVLMPP 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003s.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	95	100.0	19	AAE12295	Aae12295 Mycobacte
2	95	100.0	163	AAP91963	Aap91963 MPB-70 pr
3	95	100.0	163	AAE07053	Aae07053 Immunopro
4	95	100.0	163	AAW12045	Aaw12045 MPB 70. 4
5	95	100.0	192	AAE07235	Aae07235 Protean an
6	95	100.0	193	AAW99351	Aaw99351 M.tubercu
7	95	100.0	201	AAE06839	Aae06839 Human IL-
8	87	91.6	220	AAW17979	Aaw17979 Mycobacte
9	87	91.6	220	AAW99350	Aaw99350 M.tubercu
10	85	89.5	228	AAW14904	Aaw14904 Amino aci
11	85	89.5	228	ABE73510	Abbe73510 M vaccae
12	85	89.5	228	ABP70879	Abp70879 Mycobacte
13	85	89.5	231	AAW14905	Aaw14905 Amino aci
14	85	89.5	231	ABE73511	Abbe73511 M vaccae
15	85	89.5	231	ABP70878	Abp70878 Mycobacte
16	66	69.5	16	AAE45711	Aae45711 MPB-70 an
17	54	56.8	2157	AAE93310	Aae93310 A human h
18	54	56.8	2570	ABE82200	Abbe82200 Human CUE
19	49	51.6	180	ABG77905	Abg77905 High leve
20	47	49.5	456	AAE31800	Aae31800 Arabidops
21	47	49.5	456	AAE31799	Aae31799 Arabidops
22	47	49.5	470	AAE31798	Aae31798 Arabidops
23	46	48.4	429	ABU26510	Abu26510 Protein e
24	45	47.4	2625	AAW55887	Aaw55887 Human tel
25	45	47.4	2627	AAW61347	Aaw61347 Human cel

26	45	47.4	2627	7	ADBE63121	Adbe63121 Human Pro
27	44	46.3	18	2	AAE41874	Aae41874 Mouse OSF
28	44	46.3	811	2	AAE41867	Aae41867 Mouse OSF
29	44	46.3	811	5	AAU79824	Aau79824 OSF-2 pro
30	44	46.3	838	5	AAU79822	Aau79822 OSF-2 pro
31	43	45.3	54	4	AAW74156	Aaw74156 Human bon
32	43	45.3	54	4	AAW61390	Aaw61390 Human bra
33	43	45.3	54	4	ABG55939	Abg55939 Human liv
34	43	45.3	54	5	ABG4086	Abg4086 Human pep
35	43	45.3	61	4	ABE40870	Abbe40870 Peptide #
36	43	45.3	61	4	AAW34637	Aaw34637 Peptide #
37	43	45.3	61	4	AAW74525	Aaw74525 Human bon
38	43	45.3	61	4	AAW61729	Aaw61729 Human bra
39	43	45.3	61	4	ABG56315	Abg56315 Human liv
40	43	45.3	349	6	ABU97181	Abu97181 Enzyme po
41	43	45.3	444	4	ABE67905	Abbe67905 Drosophi
42	43	45.3	696	6	ABU37047	Abu37047 Human bre
43	43	45.3	730	6	ABP96205	Abp96205 Human mat
44	43	45.3	747	6	ABP96202	Abp96202 Human per
45	43	45.3	750	6	ABP96207	Abp96207 Human mat

ALIGNMENTS

RESULT 1
AAE12295
ID AAE12295 standard; peptide; 19 AA.
XX
AC AAE12295;
XX
DT 18-DEC-2001 (first entry)
XX
DE Mycobacterium tuberculosis (Mtb) peptide #80.
XX
KW Mycobacterium tuberculosis; Mtb peptide; antibacterial; vaccine;
XX infection; anti-Mtb immune response.
OS Mycobacterium tuberculosis.
FN WO200170774-A2.
XX
PD 27-SEP-2001.
PF 20-MAR-2001; 2001WO-US009906.
XX
PR 20-MAR-2000; 2000US-0190834P.
XX
PA (UYBR-) UNIV BROWN RES FOUND.
XX
PI Degroot AS;
XX
DR WPI; 2001-616401/71.
XX
PT New vaccine for immunizing a mammalian subject, preferably humans,
XX against infection caused by Mycobacterium tuberculosis.
PS Disclosure; Fig 4; 42pp; English.
XX

The present invention relates to Mycobacterium tuberculosis (Mtb) vaccine candidate peptides. The invention also relates to a method for identifying Mtb vaccine candidate peptides as well as vaccines comprising these candidate peptides. Vaccines of the invention and Mtb vaccine candidate peptides are useful for inducing an anti-Mtb antibody in a mammalian subject preferably human. They are used for immunising a mammalian subject, preferably humans, against infection caused by Mycobacterium tuberculosis. The present sequence is a Mtb vaccine candidate peptide

SQ Sequence 19 AA;

Query Match 100.0%; Score 95; DB 4; Length 19;

Best Local Similarity 100.0%; Pred. No. 4.3e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
Db |||||

RESULT 2

AAAP91963
ID AAP91963 standard; protein; 163 AA.

XX AC AAP91963;
XX DT 25-MAR-2003 (revised)
XX DT 22-FEB-1990 (first entry)

XX DE MPB-70 protein of Mycobacterium bovis AN5.
XX KW Mycobacterium bovis; strain AN5; MPB-70 protein.
XX OS Mycobacterium bovis.

XX PN WO8909261-A.
XX PD 05-OCT-1989.
XX PF 31-MAR-1989; 89WO-AU000143.

XX PR 31-MAR-1988; 88AU-00007550.
XX PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX PI Wood PR, Radford AJ;

XX DR WPI; 1989-309529/42.
XX DR N-PSDB; AAN91472.

XX PT Diagnosis of mycobacterium bovis infection - using antibodies specific to
XX MPB-70 protein of M bovis.

XX PS Disclosure; Fig 3; 48pp; English.
XX CC The MPB-70 protein can be detected by specific antibodies or by a cell-
XX CC mediated immune response against it, to diagnose M.bovis infection. The
XX CC protein is produced by chromatofocussing of a M.bovis AN5 culture
XX CC filtrate. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-
XX CC MAR-2003 to correct PA field.)

XX SQ Sequence 163 AA;
Query Match 100.0%; Score 95; DB 1; Length 163;
Best Local Similarity 100.0%; Pred. No. 6.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
Db |||||

RESULT 3
AAR07053
ID AAR07053 standard; protein; 163 AA.

XX AC AAR07053;
XX DT 14-JAN-1991 (first entry)

XX DE Immunoprotein MPB 70 derived from a BCG bacteria.
XX KW Bovine tuberculosis; Interleukin-2; IL-2; plasmid pT13S.
XX OS Mycobacterium bovis.

XX SQ Sequence 163 AA;
Query Match 100.0%; Score 95; DB 1; Length 163;
Best Local Similarity 100.0%; Pred. No. 6.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
Db |||||

RESULT 4
AAW12045
ID AAW12045 standard; protein; 163 AA.

XX AC AAW12045;
XX DT 02-APR-1997 (first entry)
XX DE MPB 70.

XX KW T-cell epitope; antigen; T-cell determinant; receptor; MHC protein; bird;
XX KW HIV sfg; herpes simplex virus; antigen gD2; tetanus toxoid; vaccine; HSV;
XX KW mammal; gp120; immune response; B-cell antigen.

XX OS Mycobacterium bovis.
XX PN WO9518148-A1.
XX PD 06-JUL-1995.

XX PF 28-DEC-1993; 93WO-US011703.
XX PR 28-DEC-1993; 93WO-US011703.

XX PA (CHIR-) CHIRON MIMOTOPES PTY LTD.
XX PI Geysen HM, Rodda SJ;
XX DR WPI; 1995-246333/32.

XX PT T cell epitope peptide(s) - useful for detecting exposure of a subject to
XX PT an antigen or pathogen, and in vaccines for birds and mammals.

XX PS Example 1; Page 9-10; 57pp; English.
XX CC This sequence represents the Mycobacterium bovis MPB70. This sequence was
XX CC used to create a pool of T-cell epitope peptides (see AAW12046-W12067). T
XX CC -cell epitopes (also known as T-cell determinants) are peptides (or

XX JPO2195895-A.
XX PD 02-AUG-1990.
XX PF 24-JAN-1989; 89JP-00013270.
XX PR 24-JAN-1989; 89JP-00013270.
XX PA (AJIN) AJINOMOTO KK.
XX DR WPI; 1990-278851/37.
XX DR N-PSDB; AAQ05975.
XX CC BCG bacteria derived immuno:protein MPB70 - can be used as diagnostic
XX CC agent used to determine bovine tuberculosis.
XX PS Claim 1; Page 694; 11pp; Japanese.

XX CC Immunoprotein MPB 70 encoding sequence may be incorporated into plasmid
XX CC pT13S with an N-terminal fragment of human IL-2. The plasmid may be used
XX CC to transform an expression system giving a fusion protein which may be
XX CC used as a diagnostic agent for bovine tuberculosis infection

XX SQ Sequence 163 AA;
Query Match 100.0%; Score 95; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 6.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
Db |||||

RESULT 4
AAW12045
ID AAW12045 standard; protein; 163 AA.

XX AC AAW12045;
XX DT 02-APR-1997 (first entry)
XX DE MPB 70.

XX KW T-cell epitope; antigen; T-cell determinant; receptor; MHC protein; bird;
XX KW HIV sfg; herpes simplex virus; antigen gD2; tetanus toxoid; vaccine; HSV;
XX KW mammal; gp120; immune response; B-cell antigen.

XX OS Mycobacterium bovis.
XX PN WO9518148-A1.
XX PD 06-JUL-1995.

XX PF 28-DEC-1993; 93WO-US011703.
XX PR 28-DEC-1993; 93WO-US011703.

XX PA (CHIR-) CHIRON MIMOTOPES PTY LTD.
XX PI Geysen HM, Rodda SJ;
XX DR WPI; 1995-246333/32.

XX PT T cell epitope peptide(s) - useful for detecting exposure of a subject to
XX PT an antigen or pathogen, and in vaccines for birds and mammals.

XX PS Example 1; Page 9-10; 57pp; English.
XX CC This sequence represents the Mycobacterium bovis MPB70. This sequence was
XX CC used to create a pool of T-cell epitope peptides (see AAW12046-W12067). T
XX CC -cell epitopes (also known as T-cell determinants) are peptides (or

CC regions of a protein) which bind to T-cell antigen receptors in
 CC conjugation with MHC proteins. The epitope sequences shown in AAW11953-
 CC W11976 were the most antigenic peptides obtained from pools of peptides
 CC created from the HIV sf2 gp120 (AAW11953-W11960), herpes simplex virus
 CC antigen gD2 (AAW11961-W11969), and tetanus toxoid (AAW11970-W11976). The
 CC epitopes can be used in methods for detecting exposure of a mammal or
 CC bird to an antigen, and for increasing the number of T-cells specific for
 CC an antigen. The peptides can also be used in a method for determining T-
 CC cell epitopes specific for an antigen. These methods allow for the
 CC identification of T-cell determinants. The T-cell epitope peptides can be
 CC used in a vaccine for inducing an immune response in a bird or mammal.
 CC The vaccine also contains a B-cell antigen, preferably herpes simplex
 CC virus gD2 (see AAW12068) or HIV sf2 gp120 (see AAW11977), and a carrier
 CC
 XX
 SQ Sequence 163 AA;

Query Match 100.0%; Score 95; DB 2; Length 163;
 Best Local Similarity 100.0%; Pred. No. 6.3e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
 |||||
 Db 144 GVSTANATVYMIDSVLMPP 162

RESULT 5
 AAR07235
 ID AAR07235 standard; protein; 192 AA.

AC AAR07235;
 CC
 DT 25-MAR-2003 (revised)
 DT 23-JAN-1991 (first entry)
 XX

DE Proten and and secretory region of MPB70 gene.
 XX Actinomycetes expression system; BCG; Mycobacteria.
 XX Mycobacterium bovis.

XX WO9010701-A.
 XX 20-SEP-1990.
 XX 08-MAR-1989; 89AU-00003099.
 XX 08-MAR-1989; 89AU-00003099.

XX (CSIR) COMMONWEALTH SCI & IND RES ORG.
 XX Radford A, Wood PR;
 XX WPI; 1990-305024/40.
 XX N-PSDB; AAQ06112.

XX DNA vectors contg. MPB70 gene promoter region - used as expression system
 XX for actinomycetes and related organisms, esp. mycobacterium bovis BCG.

XX Disclosure; Fig 2; 31pp; English.

XX Gene product was isolated from MPB70 gene of Mycobacterium bovis BCG.
 CC Product is encoded by a construct expressing heterologous antigenic genes
 CC for development of vaccines. Vaccines may be created which will induce a
 CC cell-mediated immune response, diagnostically distinguishable from wild
 CC type infection by Mycobacteria, eg. M.bovis, M.tuberculosis, M.lprae etc.
 CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 192 AA;

Query Match 100.0%; Score 95; DB 2; Length 192;
 Best Local Similarity 100.0%; Pred. No. 7.7e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
 |||||
 Db 173 GVSTANATVYMIDSVLMPP 191

RESULT 6

AAW99351
 ID AAW99351 standard; peptide; 193 AA.

XX AAW99351;

XX 21-MAY-1999 (first entry)

DE M.tuberculosis MPT70 protein.

XX MPT83; glycosylation motif; immunogenicity; vaccine; immunisation;
 XX mammal; infection; Mycobacterium tuberculosis; actinomycete.

XX Mycobacterium tuberculosis.

XX WO9902706-A1.

XX 21-JAN-1999.

XX 06-JUL-1998; 98WO-GB001989.

XX 07-JUL-1997; 97GB-00014242.

XX (UKAG-) UK MIN FISHERIES & FOOD.

XX Hewinson RG, Michell SL;

XX WPI; 1999-120907/10.

XX A new recombinant DNA encoding a glycosylation motif - useful to
 XX glycosylate proteins when expressed in an actinomycete host.

XX Disclosure; Page 17-18; 28pp; English.

XX This sequence corresponds to the MPT70 protein from Mycobacterium
 CC tuberculosis. The invention relates to the introduction of protein
 CC glycosylation motifs, into other proteins, especially when expressed in
 CC an actinomycete host cell, in order to improve their immunogenicity and
 CC hence their use in e.g. vaccines. The polypeptide is used to immunise a
 CC mammal against infection by M. tuberculosis

XX Sequence 193 AA;

Query Match 100.0%; Score 95; DB 2; Length 193;
 Best Local Similarity 100.0%; Pred. No. 7.8e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
 |||||
 Db 174 GVSTANATVYMIDSVLMPP 192

RESULT 7

AAR06839
 ID AAR06839 standard; protein; 201 AA.

XX AAR06839;

XX 14-JAN-1991 (first entry)

XX Human IL-2 N-terminal fragment and immunoprotein MPB 70 derived from a
 XX BCG bacteria.

XX Bovine tuberculosis; Interleukin-2; IL-2; plasmid pT13S.

XX Mycobacterium bovis.

XX JP02195895-A.

```

XX PD 02-AUG-1990.
XX PF 24-JAN-1989; 89JP-00013270.
XX PR 24-JAN-1989; 89JP-00013270.
XX PA (AJIN ) AJINOMOTO KK.
XX WPI; 1990-278851/37.
XX
XX BCG bacteria derived immuno:protein MPB70 - can be used as diagnostic
XX agent used to determine bovine tuberculosis.
XX
XX Disclosure; Fig 3; 11pp; Japanese.
XX
XX Immunoprotein MPB 70 encoding sequence may be incorporated into plasmid
XX pTl3S with an N-terminal fragment of human IL-2. The plasmid may be used
XX to transform an expression system giving a fusion protein which may be
XX used as a diagnostic agent for bovine tuberculosis infection
XX
XX SQ Sequence 201 AA;
XX
XX Query Match 100.0%; Score 95; DB 2; Length 201;
XX Best Local Similarity 100.0%; Pred. No. 8.2e-09;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GVSTANATVMIDSVLMPP 19
XX 182 GVSTANATVMIDSVLMPP 200
XX
XX RESULT 8
XX AAW17979
XX ID AAW17979 standard; protein; 220 AA.
XX AC AAW17979;
XX
XX DT 23-JUL-1997 (first entry)
XX
XX DE Mycobacterium tuberculosis mpt83 protein.
XX
XX KW Vaccine; diagnostic agent; antigen; lipoylation; glycosylation.
XX
XX OS Mycobacterium tuberculosis.
XX
XX PH Key Location/Qualifiers
XX FT Peptide 1..21
XX FT /label= Secretion_signal
XX FT Misc-difference 22..26
XX FT /label= Lipoylation_motif
XX FT Misc-difference 35..57
XX FT /label= Glycosylation_motif
XX
XX PN WO9708322-A1.
XX
XX PD 06-MAR-1997.
XX
XX PF 19-AUG-1996; 96WO-GB002015.
XX
XX PR 25-AUG-1995; 95GB-00017494.
XX
XX PA (UKAG-) UK MIN FISHERIES & FOOD.
XX
XX PI Hewinson RG, Jacobs WR;
XX
XX WPI; 1997-179279/16.
XX
XX N-PSDB; AAT70155.
XX
XX New isolated mpt83 gene from Mycobacterium tuberculosis - used to develop
XX prods. for use as vaccines or as diagnostic agents.
XX
XX Claim 1; Page 23-24; 40pp; English.

```

```

XX The gene mpt83 encodes the 25 kDa antigen of Mycobacterium tuberculosis.
XX The present sequence represents a 220 amino acid mpt83 gene product. The
XX protein can be used in vaccines to protect against Mycobacterium
XX tuberculosis infection or for antibody production. The antibodies are
XX useful as diagnostic agents, to detect M. tuberculosis infection in a
XX sample. The mpt83 gene promoter, glycosylation, lipoylation and secretion
XX sequences can be used in recombinant DNA expression systems for use in
XX the transformation of cells e.g. to produce glycosylated or lipoylated
XX products which can be regulated by macrophage factors
XX
XX SQ Sequence 220 AA;
XX
XX Query Match 91.6%; Score 87; DB 2; Length 220;
XX Best Local Similarity 89.5%; Pred. No. 2.6e-07;
XX Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 GVSTANATVMIDSVLMPP 19
XX 200 GVHTANATVMIDTVLMPP 218
XX
XX RESULT 9
XX AAW99350
XX ID AAW99350 standard; peptide; 220 AA.
XX AC AAW99350;
XX
XX DT 21-MAY-1999 (first entry)
XX
XX DE M.tuberculosis MPT83 protein.
XX
XX KW MPT83; glycosylation motif; immunogenicity; vaccine; immunisation;
XX mammal; infection; Mycobacterium tuberculosis; actinomycete.
XX
XX OS Mycobacterium tuberculosis.
XX
XX PN WO9902706-A1.
XX
XX PD 21-JAN-1999.
XX
XX PF 06-JUL-1998; 98WO-GB001989.
XX
XX PR 07-JUL-1997; 97GB-00014242.
XX
XX PA (UKAG-) UK MIN FISHERIES & FOOD.
XX
XX PI Hewinson RG, Michell SL;
XX
XX WPI; 1999-120907/10.
XX
XX A new recombinant DNA encoding a glycosylation motif - useful to
XX glycosylate proteins when expressed in an actinomycete host.
XX
XX Disclosure; Page 16-17; 28pp; English.
XX
XX This sequence corresponds to the MPT83 protein from Mycobacterium
XX tuberculosis. The invention relates to the introduction of protein
XX glycosylation motifs, into other proteins, especially when expressed in
XX an actinomycete host cell, in order to improve their immunogenicity and
XX hence their use in e.g. vaccines. The polypeptide is used to immunise a
XX mammal against infection by M. tuberculosis
XX
XX SQ Sequence 220 AA;
XX
XX Query Match 91.6%; Score 87; DB 2; Length 220;
XX Best Local Similarity 89.5%; Pred. No. 2.6e-07;
XX Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 GVSTANATVMIDSVLMPP 19
XX 200 GVHTANATVMIDTVLMPP 218
XX

```

RESULT 11


```

DR WPI; 2002-138361/18.
DR N-PSDB; ABL36273.
XX
XX Inhibiting skin inflammation associated with skin disorder e.g.
XX psoriasis, by administering composition comprising delipidated and
XX deglycolipidated Mycobacterium vaccae cells or Mycobacterium vaccae
XX culture filtrate.
XX
XX Example 5; Col 157-160; 116pp; English.
XX
XX The present invention relates to a method of inhibiting skin inflammation
XX associated with a skin disorder selected from psoriasis, atopic
XX dermatitis and allergic contact dermatitis, which involves administering
XX a composition containing delipidated and deglycolipidated Mycobacterium
XX vaccae cells or M. vaccae culture filtrate. The skin disorder to be
XX treated may also include alopecia areata, and skin cancers such as basal
XX cell carcinoma, squamous cell carcinoma and melanoma. The composition
XX acts by inhibiting the Th2 immune response. The present sequence is a
XX protein described in the exemplification of the invention
XX
XX Sequence 231 AA;
SQ
Query Match 89.5%; Score 85; DB 5; Length 231;
Best Local Similarity 84.2%; Pred. No. 6.4e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
||| ||||| ||||| |||||
Db 196 GVQTANATVYLIDTVLMPP 214

RESULT 15
ABF70878
ID ABP70878 standard; protein; 231 AA.
XX
AC ABP70878;
XX
DT 26-AUG-2003 (first entry)
XX
DE Mycobacterium vaccae antigen GV-1/70, SEQ ID 27.
XX
XX Immunosuppressive; neuroprotective; antirheumatic; antiarthritic;
XX antidiabetic; antipsoriatic; dermatological; anti-inflammatory;
XX immune response; Notch signalling pathway; autoimmune disorder;
XX toll-like receptor signalling pathway; antigen; allergy; graft rejection;
XX apoptotic cell death; cell proliferation.
XX
XX Mycobacterium vaccae.
XX
XX Key Location/Qualifiers
FH Misc-difference 1..7 /notes="Encoded by CCC"
FT Misc-difference 9..10 /notes="Encoded by AAC TAA AAC"
FT Misc-difference 216..217 /notes="Encoded by GCG TAG CCG"
XX
XX WO2003013595-A1.
XX
XX 20-FEB-2003.
XX
XX 26-JUL-2002; 2002WO-NZ000135.
XX
XX 26-JUL-2001; 2001US-0308446P.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX Watson JD, Tan PLJ, Abernethy N;
XX
XX WPI; 2003-239567/23.
XX
XX N-PSDB; ACC42518.
XX
XX Methods for modulating immune responses by modulating the Notch signaling

```

```

PT and Toll-like receptor signaling pathways, and treating e.g. autoimmune
PT disorders.
XX
XX Claim 7; Page 114; 136pp; English.
XX
XX The present invention relates to methods for modulating immune responses
XX by modulating the Notch signalling and Toll-like receptor signalling
XX pathways using compositions comprising mycobacteria antigens (ACC42518-
XX ACC42543 and ABP70878-ABP70903). The methods are useful for modulating
XX immune responses and treating e.g. autoimmune disorders (such as multiple
XX sclerosis, rheumatoid arthritis, Type I diabetes mellitus, psoriasis
XX systemic lupus erythematosus, scleroderma), allergic disease and graft
XX rejection and also disorders characterised by undesired apoptotic cell
XX death or undesired cell proliferation
XX
XX Sequence 231 AA;
SQ
Query Match 89.5%; Score 85; DB 6; Length 231;
Best Local Similarity 84.2%; Pred. No. 6.4e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
||| ||||| ||||| |||||
Db 196 GVQTANATVYLIDTVLMPP 214

Search completed: March 10, 2004, 12:05:40
Job time : 47.1859 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 12:11:07 ; Search time 23.8718 Seconds
(without alignments)
168.061 Million cell updates/sec

Title: US-10-044-703-80

Perfect score: 95

Sequence: 1 GVSTANATVMIDSVLMP 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
1	95	100.0	19	9	US-09-813-333-80	Sequence 80, Appl
2	95	100.0	19	13	US-10-044-703-80	Sequence 80, Appl
3	85	89.5	228	10	US-09-880-505-147	Sequence 147, App
4	85	89.5	228	13	US-10-051-643-147	Sequence 147, App
5	85	89.5	228	14	US-10-205-979-28	Sequence 28, Appl
6	85	89.5	231	10	US-09-880-505-152	Sequence 152, App
7	85	89.5	231	13	US-10-051-643-152	Sequence 152, App
8	85	89.5	231	14	US-10-205-979-27	Sequence 27, Appl
9	54	56.8	2212	15	US-10-028-248A-43	Sequence 43, Appl
10	54	56.8	2212	15	US-10-107-782-43	Sequence 43, Appl
11	54	56.8	2570	15	US-10-028-248A-42	Sequence 42, Appl
12	54	56.8	2570	15	US-10-107-782-42	Sequence 42, Appl
13	49	51.6	180	13	US-10-047-260-18	Sequence 18, Appl
14	46	48.4	149	15	US-10-028-248A-44	Sequence 44, Appl
15	46	48.4	149	15	US-10-028-248A-45	Sequence 45, Appl

16	46	48.4	149	15	US-10-107-782-44	Sequence 44, Appl
17	46	48.4	149	15	US-10-107-782-45	Sequence 45, Appl
18	45	47.4	2630	15	US-10-334-143-41	Sequence 41, Appl
19	43	45.3	54	9	US-09-864-761-44557	Sequence 44557, A
20	43	45.3	61	9	US-09-864-761-4457	Sequence 44726, A
21	43	45.3	349	14	US-10-166-225A-70	Sequence 70, Appl
22	43	45.3	696	14	US-10-176-847-46	Sequence 46, Appl
23	43	45.3	730	14	US-10-217-371-6	Sequence 6, Appl
24	43	45.3	750	14	US-10-217-371-14	Sequence 14, Appl
25	43	45.3	751	14	US-10-217-371-4	Sequence 4, Appl
26	43	45.3	758	14	US-10-217-371-10	Sequence 10, Appl
27	43	45.3	771	14	US-10-217-371-12	Sequence 12, Appl
28	43	45.3	779	14	US-10-217-371-8	Sequence 8, Appl
29	43	45.3	779	14	US-10-171-311-180	Sequence 180, Appl
30	43	45.3	779	14	US-10-301-822-151	Sequence 151, Appl
31	43	45.3	790	9	US-09-925-301-1313	Sequence 1313, App
32	43	45.3	836	14	US-10-217-371-2	Sequence 2, Appl
33	43	45.3	836	14	US-10-171-311-178	Sequence 178, App
34	43	45.3	836	14	US-10-177-293-342	Sequence 342, App
35	43	45.3	836	14	US-10-301-822-149	Sequence 149, App
36	43	45.3	836	14	US-10-204-752-33	Sequence 33, Appl
37	43	45.3	836	15	US-10-295-027-36	Sequence 36, Appl
38	43	45.3	836	15	US-10-173-999-66	Sequence 66, Appl
39	41	43.2	34	9	US-09-864-761-44399	Sequence 44399, A
40	41	43.2	158	15	US-10-108-260A-4106	Sequence 4106, App
41	41	43.2	437	14	US-10-232-655-1	Sequence 1, Appl
42	41	43.2	683	9	US-09-922-217-122	Sequence 122, App
43	41	43.2	683	9	US-09-833-263-122	Sequence 122, App
44	41	43.2	683	13	US-10-025-380-122	Sequence 122, App
45	41	43.2	683	14	US-10-301-822-201	Sequence 201, App

ALIGNMENTS

RESULT 1

US-09-813-333-80
; Sequence 80, Application US/09813333
; Patent No. US2002019160A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/09/813,333
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/190,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 80
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-333-80

Query Match 100.0%; Score 95; DB 9; Length 19;

Best Local Similarity 100.0%; Pred. No. 2.8e-09;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVSTANATVMIDSVLMP 19

Db 1 GVSTANATVMIDSVLMP 19

RESULT 2

US-10-044-703-80
; Sequence 80, Application US/10044703
; Publication No. US20020192233A1
; GENERAL INFORMATION:
; APPLICANT: Degroot, Anne S
; TITLE OF INVENTION: Human T Cell Response to MHC-Binding Motif Clusters
; FILE REFERENCE: 17999-004 US
; CURRENT APPLICATION NUMBER: US/10/044,703

; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 60/130,834
; PRIOR FILING DATE: 2000-03-20
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 80
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-044-703-80

Query Match 100.0%; Score 95; DB 13; Length 19;
Best Local Similarity 100.0%; Pred. No. 2.8e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVSTANATVYMIDSVLMP 19
Db 1 GVSTANATVYMIDSVLMP 19

RESULT 3

US-09-880-505-147
; Sequence 147, Application US/09880505
; Publication No. US20030007976A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
; FILE REFERENCE: 11000.1007c2
; CURRENT APPLICATION NUMBER: US/09/880,505
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 09/324,542
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: US 08/997,080
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 147
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-09-880-505-147

Query Match 89.5%; Score 85; DB 10; Length 228;
Best Local Similarity 84.2%; Pred. No. 2.4e-06;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GVSTANATVYMIDSVLMP 19
Db 208 GVQTANATVYLDITVLMPP 226

RESULT 4

US-10-051-643-147
; Sequence 147, Application US/10051643
; Publication No. US20020197265A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Diseases of the Respiratory
; TITLE OF INVENTION: System using Mycobacterium vaccae
; FILE REFERENCE: 11000.1008c2
; CURRENT APPLICATION NUMBER: US/10/051,643
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US09/156,181
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: US 08/996,624
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 147
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-10-051-643-147

Query Match 89.5%; Score 85; DB 13; Length 228;
Best Local Similarity 84.2%; Pred. No. 2.4e-06;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GVSTANATVYMIDSVLMP 19
Db 208 GVQTANATVYLDITVLMPP 226

RESULT 5

US-10-205-979-28
; Sequence 28, Application US/10205979
; Publication No. US20030147861A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; APPLICANT: Abernethy, Nevin
; TITLE OF INVENTION: Compounds and Methods for the Modulation
; TITLE OF INVENTION: of Immune Responses
; FILE REFERENCE: 11000.1063U
; CURRENT APPLICATION NUMBER: US/10/205,979
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/308,446
; PRIOR FILING DATE: 2001-07-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-10-205-979-28

Query Match 89.5%; Score 85; DB 14; Length 228;
Best Local Similarity 84.2%; Pred. No. 2.4e-06;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GVSTANATVYMIDSVLMP 19
Db 208 GVQTANATVYLDITVLMPP 226

RESULT 6

US-09-880-505-152
; Sequence 152, Application US/09880505
; Publication No. US20030007976A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
; FILE REFERENCE: 11000.1007c2
; CURRENT APPLICATION NUMBER: US/09/880,505
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 09/324,542
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: US 08/997,080
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 152
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-09-880-505-152

Query Match 89.5%; Score 85; DB 10; Length 231;

Best Local Similarity 84.2%; Pred. No. 2.5e-06;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
|||:|||||:|||||
Db 196 GVQTANATVYLDIVLMPP 214

RESULT 7
US-10-051-643-152
; Sequence 152, Application US/10051643
; Publication No. US20020197265A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Diseases of the Respiratory
; FILE REFERENCE: 11000.1008c2
; CURRENT APPLICATION NUMBER: US/10/051,643
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US09/156,181
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: US 08/996,624
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 152
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-10-051-643-152

Query Match 89.5%; Score 85; DB 13; Length 231;
Best Local Similarity 84.2%; Pred. No. 2.5e-06;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
|||:|||||:|||||
Db 196 GVQTANATVYLDIVLMPP 214

RESULT 8
US-10-205-979-27
; Sequence 27, Application US/10205979
; Publication No. US20030147861A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; APPLICANT: Abernethy, Nevin
; TITLE OF INVENTION: Compounds and Methods for the Modulation
; TITLE OF INVENTION: of Immune Responses
; FILE REFERENCE: 11000.1063U
; CURRENT APPLICATION NUMBER: US/10/205,979
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/308,446
; PRIOR FILING DATE: 2001-07-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-10-205-979-27

Query Match 89.5%; Score 85; DB 14; Length 231;
Best Local Similarity 84.2%; Pred. No. 2.5e-06;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVSTANATVYMIDSVLMPP 19
|||:|||||:|||||
Db 196 GVQTANATVYLDIVLMPP 214

RESULT 9
US-10-028-248A-43
; Sequence 43, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shimkete, Richard
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malyankar, Uziel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glennda
; APPLICANT: Zernusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
; FILE REFERENCE: 21402-223
; CURRENT APPLICATION NUMBER: US/10/028,248A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 2212
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-028-248A-43

Query Match 56.8%; Score 54; DB 15; Length 2212;
Best Local Similarity 44.4%; Pred. No. 5.8;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSTANATVYMIDSVLMPP 19
|||:|||||:|||||
Db 302 VMAANGVIHMDGILPP 319

RESULT 10
US-10-107-782-43
; Sequence 43, Application US/10107782
; Publication No. US20040018970A1
; GENERAL INFORMATION:
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Casman, Stacie
; APPLICANT: Colman, Steve,
; APPLICANT: Edinger, Shlomit,

```

; APPLICANT: Gangolli, Esha,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Li, Li,
; APPLICANT: Liu, Xiaohong,
; APPLICANT: Malyankar, Uriel,
; APPLICANT: Miller, Charles,
; APPLICANT: Millet, Isabelle,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Sciore, Paul,
; APPLICANT: Shenoy, Suresh,
; APPLICANT: Shinkets, Richard,
; APPLICANT: Si, Jingsheng,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Stone, David,
; APPLICANT: Taupier, Raymond, jr.,
; APPLICANT: Tchernev, Velizar,
; APPLICANT: Vernet, Corine,
; APPLICANT: Zerhusen, Brian
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 21402-222CIP
; CURRENT APPLICATION NUMBER: US/10/107,782
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 10/028,248
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256,619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262,959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272,408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285,189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308,039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311,266
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/279,344
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: Curasequest version 0.1
; SEQ ID NO 43
; LENGTH: 2212
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-107-782-43

```

```

Query Match      56.8%; Score 54; DB 15; Length 2212;
Best Local Similarity 44.4%; Pred. No. 5.8;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy      2 VSTANATVYMIDSVLMPP 19
Db      302 VMAANGVIHMLDGIILPP 319

```

```

RESULT 11
US-10-028-248A-42
; Sequence 42, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shinkets, Richard
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malyankar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li

```

```

; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glenda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/2629959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42
; LENGTH: 2570
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-028-248A-42

```

```

Query Match      56.8%; Score 54; DB 15; Length 2570;
Best Local Similarity 44.4%; Pred. No. 6.9;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy      2 VSTANATVYMIDSVLMPP 19
Db      627 VMAANGVIHMLDGIILPP 644

```

```

RESULT 12
US-10-107-782-42
; Sequence 42, Application US/10107782
; Publication No. US20040018970A1
; GENERAL INFORMATION:
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Casman, Stacie
; APPLICANT: Colman, Steve,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Gangolli, Esha,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Li, Li,
; APPLICANT: Liu, Xiaohong,
; APPLICANT: Malyankar, Uriel,
; APPLICANT: Miller, Charles,
; APPLICANT: Millet, Isabelle,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Sciore, Paul,
; APPLICANT: Shenoy, Suresh,
; APPLICANT: Shinkets, Richard,
; APPLICANT: Si, Jingsheng,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Stone, David,
; APPLICANT: Taupier, Raymond, jr.,

```

```

; APPLICANT: Tchernev, Velizar,
; APPLICANT: Vernet, Corine,
; APPLICANT: Zerhusen, Brian
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES AND METHODS OF USE THEREOF
; CURRENT APPLICATION NUMBER: US/10/107,782
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 10/028,248
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256,619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262,959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272,408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285,189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308,039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311,266
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/279,344
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 42
; LENGTH: 2570
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-107-782-42

```

```

Query Match      56.8%; Score 54; DB 15; Length 2570;
Best Local Similarity 44.4%; Pred. No. 6.9;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      2 VSTANATVYMIDSVLMPP 19
DB      627 VVAANGVIHMLDGIILPP 644

```

```

RESULT 13
US-10-047-260-18
; Sequence 18, Application US/10047260
; Publication No. US20020164706A1
; GENERAL INFORMATION:
; APPLICANT: Huang, Lisa
; APPLICANT: McClusky, Michael
; APPLICANT: Larossa, Robert
; TITLE OF INVENTION: HIGH LEVEL PROMOTERS FROM CYANOBACTERIA
; FILE REFERENCE: CIL715 US NA
; CURRENT APPLICATION NUMBER: US/10/047,260
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/264,925
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Synechocystis sp. strain PCC6803
US-10-047-260-18

```

```

Query Match      51.6%; Score 49; DB 13; Length 180;
Best Local Similarity 38.9%; Pred. No. 2.3;
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      2 VSTANATVYMIDSVLMPP 19
DB      162 VDAANGVIHVIDQVILPP 179

```

```

RESULT 14
US-10-028-248A-44

```

```

; Sequence 44, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shinketsu, Richard
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malyankar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glennda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: No. US20030235882A1el Nucleic Acids and Polypeptides and Methods
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fasciclin
; OTHER INFORMATION: Domain Sequence
US-10-028-248A-44

```

```

Query Match      48.4%; Score 46; DB 15; Length 149;
Best Local Similarity 41.2%; Pred. No. 5.8;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      2 VSTANATVYMIDSVLMPP 18
DB      133 IETTINGVIHVIDRVLLP 149

```

```

RESULT 15
US-10-028-248A-45
; Sequence 45, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shinketsu, Richard
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie

```

```
; APPLICANT: Malyankar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Baha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glennda
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rotherberg, Mark
; TITLE OF INVENTION: No. US2003023582A1el Nucleic Acids and Polypeptides and Methods
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fasciclin
; OTHER INFORMATION: domain sequence
US-10-028-248a-45

Query Match      48.4%; Score 46; DB 15; Length 149;
Best Local Similarity 41.2%; Pred. No. 5.8;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY      2 VSTANATVYMDSVLP 18
Db      133 IETNGVHVIDRVL 149
      :||:|||||:|:|
      :||:|||||:|:|

Search completed: March 10, 2004, 12:41:32
Job time : 23.8718 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:58:01 ; Search time 10.2308 Seconds
(without alignments)
178.641 Million cell updates/sec

Title: US-10-044-703-80

Perfect score: 95

Sequence: 1 GVSTANATVMIDSVLMPP 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*
1: P1:1:*
2: P1:2:*
3: P1:3:*
4: P1:4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	95	100.0	193	2 A48320	major secreted imm
2	95	100.0	193	2 A37195	major secreted pro
3	95	100.0	193	2 F70923	probable mpt70 pro
4	97	91.6	220	2 D70923	probable mpt83 pro
5	65	68.4	220	2 T34986	probable secreted
6	57	60.0	623	2 F75223	osteoblast specifi
7	52	54.7	160	2 F95334	Nex18 Symbolicall
8	52	54.7	178	2 E87300	conserved hypoteth
9	51	53.7	385	2 S49111	peptide synthetase
10	49	51.6	180	2 S76811	transforming growt
11	47.5	50.0	406	2 D90492	hypothetical prote
12	47	49.5	470	2 T02319	hypothetical prote
13	46	48.4	220	2 AH2463	hypothetical prote
14	46	48.4	429	2 E81354	UDP-N-acetylglucos
15	45	47.4	7463	2 T36248	CDA peptide synthe
16	44	46.3	811	2 S36109	osteoblast-specifi
17	43	45.3	185	2 AE2634	conserved hypoteth
18	43	45.3	217	2 C97416	hypothetical 18.5K
19	43	45.3	349	2 H69162	conserved hypoteth
20	43	45.3	402	2 T04348	endosporm specific
21	43	45.3	779	2 S36111	osteoblast-specifi
22	43	45.3	836	2 S36110	osteoblast-specifi
23	43	45.3	3670	2 T36249	CDA peptide synthe
24	42	44.2	141	2 AF2417	hypothetical prote
25	42	44.2	240	2 C84590	probable surface p
26	42	44.2	261	2 AF2280	hypothetical prote
27	42	44.2	289	2 S75197	ABC-type transport
28	42	44.2	492	2 C96521	protein F21D18.18
29	42	44.2	558	2 AE2033	hypothetical prote

30 42 44.2 814 1 C40618 fimbrial outer mem
31 42 44.2 838 2 AC1064 outer membrane fim
32 42 44.2 1311 2 T08986 hypotheical prote
33 42 44.2 2514 1 MNWVN2 nonstructural poly
34 42 44.2 2609 2 T44808 mycosubtilin synth
35 42 44.2 5369 2 T44807 mycosubtilin synth
36 41 43.2 140 2 AD2423 hypotheical prote
37 41 43.2 683 2 I52996 transforming growt
38 41 43.2 1305 2 T31096 cyclin G-associate
39 40.5 42.6 1622 2 JE0378 DNA (cytosine-5)-
40 40 42.1 205 2 S49791 probable membrane
41 40 42.1 235 2 T47286 hypotheical prote
42 40 42.1 256 2 T11233 H+-transporting tw
43 40 42.1 275 2 H96981 sugar permease [im
44 40 42.1 327 2 AC0285 CorA-like Mg2+ tra
45 40 42.1 362 2 S77024 hypotheical prote

ALIGNMENTS

RESULT 1

A48320
major secreted immunogenic protein MPB70 precursor - Mycobacterium bovis (strain BCG)
C:Species: Mycobacterium bovis
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 10-Mar-1994
C:Accession: A48320
R:Terazaka, K.; Yamaguchi, R.; Matsuo, K.; Yamazaki, A.; Nagai, S.; Yamada, T.
FEMS Microbiol. Lett. 58, 273-276, 1989

A:Title: Complete nucleotide sequence of immunogenic protein MPB70 from Mycobacterium ;
A:Reference number: A48320
A:Accession: A48320
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-193 <TER>
A:Cross-references: GB:X17086

Query Match 100.0%; Score 95; DB 2; Length 193;

Best Local Similarity 100.0%; Pred. No. 6.5e-09;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVSTANATVMIDSVLMPP 19
Db 174 GVSTANATVMIDSVLMPP 192

RESULT 2

A37195
major secreted protein MPB70 precursor - Mycobacterium bovis
C:Species: Mycobacterium bovis
C:Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 22-Oct-1999

C:Accession: A37195
R:Radford, A.J.; Wood, P.R.; Billman-Jacobe, H.; Geyssen, H.M.; Mason, T.J.; Tribbick, C.
J. Gen. Microbiol. 136, 265-272, 1990
A:Title: Epitope mapping of the Mycobacterium bovis secretory protein MPB70 using over;

A:Reference number: A37195; PMID:1691265
A:Accession: A37195
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-193 <RAD>

A:Cross-references: GB:M33916; GB:M25386; MID:g149975; FIDN:AAA25366.1; PID:g149976

Query Match 100.0%; Score 95; DB 2; Length 193;

Best Local Similarity 100.0%; Pred. No. 6.5e-09;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVSTANATVMIDSVLMPP 19
Db 174 GVSTANATVMIDSVLMPP 192

RESULT 3

F70923

probable mpt70 protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: F70923
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: F70923
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-193 <COL>
A: Cross-references: GB:Z74024; GB:AL123456; NID: G3250700; PID: CAA98373.1; PID: e354630;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: mpt70

Query Match 100.0%; Score 95; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 6.5e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVSTANATVMIDSVLMPP 19
|||
Db 174 GVSTANATVMIDSVLMPP 192
|||

RESULT 4
D70923
probable mpt83 protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: D70923
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: D70923
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-220 <COL>
A: Cross-references: GB:Z74024; GB:AL123456; NID: G3250700; PID: CAA98350.1; PID: e1301024;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: mpt83

Query Match 91.6%; Score 87; DB 2; Length 220;
Best Local Similarity 89.5%; Pred. No. 1.9e-07;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVSTANATVMIDSVLMPP 19
|||
Db 200 GVHTANATVMIDTVLMPP 218
|||

RESULT 5
T34986
probable secreted protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T34986
R: Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1998
A: Reference number: Z21550
A: Accession: T34986
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-220 <COL>

probable mpt83 protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: F70923
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: F70923
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-193 <COL>
A: Cross-references: GB:Z74024; GB:AL123456; NID: G3250700; PID: CAA98373.1; PID: e354630;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: mpt70

Query Match 100.0%; Score 95; DB 2; Length 193;
Best Local Similarity 100.0%; Pred. No. 6.5e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVSTANATVMIDSVLMPP 19
|||
Db 174 GVSTANATVMIDSVLMPP 192
|||

RESULT 4
D70923
probable mpt83 protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: D70923
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: D70923
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-220 <COL>
A: Cross-references: GB:Z74024; GB:AL123456; NID: G3250700; PID: CAA98350.1; PID: e1301024;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: mpt83

Query Match 91.6%; Score 87; DB 2; Length 220;
Best Local Similarity 89.5%; Pred. No. 1.9e-07;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVSTANATVMIDSVLMPP 19
|||
Db 200 GVHTANATVMIDTVLMPP 218
|||

RESULT 5
T34986
probable secreted protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T34986
R: Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1998
A: Reference number: Z21550
A: Accession: T34986
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-220 <COL>

A: Cross-references: EMBL:AL031182; PIDN: CAA20163.1; GSPDB: GNO0070; SCODB: SC4A2.11c
A: Experimental source: strain A3(2)
C: Genetics:
A: Gene: SCODB: SC4A2.11c

Query Match 68.4%; Score 65; DB 2; Length 220;
Best Local Similarity 61.1%; Pred. No. 0.0013;
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 VSTANATVMIDSVLMPP 19
:|||||:|||||
Db 191 IPTKNATVIVDGVLMPP 208
:|||||:|||||

RESULT 6
F75523
osteoblast specific factor 2-related protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75523
R: White, O.; Eisen, J.A.; Heidelberg, J.P.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A: Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A: Reference number: A75250; MUID: 20036896; PMID: 10567266
A: Accession: F75523
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-623 <WHI>
A: Cross-references: GB:AE001900; GB:AE000513; NID: G6458079; PIDN: AAF09579.1; PID: G645808
A: Experimental source: strain R1
C: Genetics:
A: Gene: DR0399
A: Map position: 1

Query Match 60.0%; Score 57; DB 2; Length 623;
Best Local Similarity 50.0%; Pred. No. 0.1;
Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 VSTANATVMIDSVLMPP 19
:|||||:|||||
Db 603 ITAGTSTVYIDTVLMPP 620
:|||||:|||||

RESULT 7
F95334
Nex18 Symbiotically induced conserved protein nex18 [imported] - Sinorhizobium meliloti
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: F95334
R: Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows L.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A: Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
A: Reference number: A95262; MUID: 21396509; PMID: 11481432
A: Accession: F95334
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-160 <KUR>
A: Cross-references: GB:AE006469; PIDN: AAK65240.1; PID: G14523690; GSPDB: GNO0165
A: Experimental source: strain 1021, megaplasmid pSymA
R: Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A: Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kise, B.; Komp, C.; Lelaure, hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A: Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A: Reference number: A96039; MUID: 21368234; PMID: 11474104
A: Contents: annotation
C: Genetics:
A: Gene: nex18

A.Accession: T02319
 A.Accession: translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 1-470 <EOU>
 A.Cross-references: EMBL:AC004481; NID:g3337347; PID:g3337362
 A.Experimental source: cultivar Columbia
 R.Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.
 M.; Koo, H.; Moffat, K.C.; Cronin, L.A.; Shen, M.; VanAken, S.
 eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.

Nature 402, 761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84755
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1,470 <STO>
A;Cross-references: GB:AE002093; NID:G3337362; PIDN:AAC27407.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g34310; P13P17.15
A;Map position: 2
A;Introns: 51/3; 95/3; 124/3; 173/3; 217/3; 247/3; 290/1

Query Match 48.4%; Score 46; DB 2; Length 429;
Best Local Similarity 56.2%; Pred. No. 5.6;
Matches 9; Conservative 4; Mismatches 3; Indels

Query Match 47.4%; Score 45; DB 2; Length 7463;
Best Local Similarity 44.4%; Pred. No. 2.1e+02;
Matches 8; Conservative 4; Mismatches 6; Indels

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:51:36 ; Search time 5.84615 Seconds
(without alignments)

169.228 Million cell updates/sec

Title: US-10-044-703-80

Perfect score: 95

Sequence: 1 GVSTANATVMIDSLVMP 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	95	100.0	193	1 MP70 MYCTU	Q50769 mycobacteri
2	87	91.6	220	1 MP83 MYCTU	Q10790 mycobacteri
3	49	51.6	180	1 YE83 SYN3	P74615 synchocyst
4	47	49.5	683	1 BGR3 RABIT	Q95215 cryctoclagus
5	43	45.3	349	1 ID12 METH	O26154 methanobact
6	42	44.2	814	1 SEFC SALEN	P33388 salmonella
7	42	44.2	2514	1 POLN ONNVG	P13886 o'nyong-nyo
8	41	43.2	683	1 BGR3 HUMAN	Q15882 homo sapien
9	41	43.2	1305	1 GAK RAT	P97874 rattus norv
10	41	43.2	1311	1 GAK HUMAN	Q14976 homo sapien
11	40.5	42.6	1620	1 DNMI MOUSE	P13864 mus musculu
12	40.5	42.6	1622	1 DNMI RAT	Q92330 rattus norv
13	40	42.1	205	1 Y119 YEAST	P40500 saccharomyc
14	40	42.1	404	1 ISGS METTE	P57795 methanosarc
15	40	42.1	653	1 ST34 ARATH	Q91w86 arabidopsis
16	40	42.1	683	1 BGR3 PIG	O11780 sus scrofa
17	40	42.1	874	1 SEGA BUCAP	Q8K9u3 buchnera ap
18	40	42.1	1173	1 GIGA OKISA	Q9aw17 oryza sativ
19	40	42.1	2505	1 FAS RAT	P12785 rattus norv
20	39.5	41.6	1277	1 CAML FUGRU	Q98902 figu rubrip
21	39	41.1	110	1 FLIE RALSO	Q8xss8 ralsstona s
22	39	41.1	252	1 BTG3 HUMAN	Q14201 homo sapien
23	39	41.1	318	1 NSR LACLA	P23648 lactococcus
24	39	41.1	342	1 UPDI CAEEL	Q19584 caenorhabdi
25	39	41.1	607	1 SYRC YEAST	Q05506 saccharomyc
26	39	41.1	623	1 CYSN XYLFA	Q9pd78 x cyen/cysc
27	39	41.1	655	1 YS44 CAEEL	Q09372 caenorhabdi
28	39	41.1	1193	1 LMG2 HUMAN	Q13753 homo sapien
29	39	41.1	3674	1 SPCR HUMAN	Q9nrc5 homo sapien
30	38	40.0	160	1 RRA1 VIBPA	Q87sd2 vibrio para
31	38	40.0	178	1 PYR2 THERIN	Q8r9r3 thermocaneer
32	38	40.0	309	1 OAS1 HUMAN	Q9ngb4 homo sapien
33	38	40.0	364	1 EZB1 PYRHO	O58433 pyrococcus

34	38	40.0	402	1 Y718 METJA	Q58128 methanococc
35	38	40.0	413	1 AIT4 MOUSE	Q00897 mus musculu
36	38	40.0	442	1 TRME MYCPN	P75104 mycoplasma
37	38	40.0	472	1 VL2 HPV34	P36758 human papil
38	38	40.0	539	1 GSHE ARATH	P46416 arabidopsis
39	38	40.0	633	1 NODQ RHISB	O07309 r nodq bifu
40	38	40.0	640	1 MT30 YEAST	P39014 saccharomyc
41	38	40.0	834	1 SHTA DROME	P28285 drosophila
42	38	40.0	1455	1 CPSA DROME	Q9v726 drosophila
43	38	40.0	1716	1 RPAL RAT	O54889 rattus norv
44	38	40.0	1717	1 RPAL MOUSE	Q35134 mus musculu
45	38	40.0	4447	1 PKSK_BACSU	P40803 bacillus su

ALIGNMENTS

RESULT 1
MP70 MYCTU
ID MP70 MYCTU STANDARD; PRT; 193 AA.
AC Q50769; Q48934; Q48946; Q48947; Q48948; Q50656;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Immunogenic protein MP70 precursor.
GN MP70 OR MP87 OR RV2875 OR MT2943 OR MTCY274.06 OR MB2900.
OS Mycobacterium tuberculosis, and
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=95176187; PubMed=7871388;
RA Matsumoto S., Matsuo T., Ohara N., Hotokezaka H., Naitoh M.,
Minami J., Yamada T.;
RT "Cloning and sequencing of a unique antigen MPT70 from Mycobacterium
tuberculosis H37Rv and expression in BCG using E. coli-mycobacteria
shuttle vector";
RT Scand. J. Immunol. 41:281-287(1995).
RL [2]
RN SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
Davies R., Brown D., Brown D., Chillingworth T., Connor R.,
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Holroyd S.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence";
RT Nature 393:537-544(1998).
RL [3]
RN SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=2206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
Bisshai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains";
RT J. Bacteriol. 184:5479-5490(2002).
RL [4]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP SPECIES=M.bovis; STRAIN=BCG / Pasteur, and BCG / Tokyo;
RX MEDLINE=89306542; PubMed=2663636;
RA Terasaka K., Yamaguchi R., Matsuo K., Yamazaki A., Nagai S.,

RA Yamada T.;
 RT "Complete nucleotide sequence of immunogenic protein MPB70 from
 RL Mycobacterium bovis BCG";
 RL FEMS Microbiol. Lett. 49:273-276(1989).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=BCG / Pasteur, and BCG / Tokyo;
 RX MEDLINE=90218009; PubMed=1691265;
 RA Radford A., Wood P., Billman-Jacobe H., Geysen H., Mason T.,
 RA Tribbick G.;
 RA "Epitope mapping of the Mycobacterium bovis secretory protein MPB70
 RT using overlapping peptide analysis";
 RL J. Gen. Microbiol. 136:265-272(1990).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=BCG / Pasteur, and BCG / Tokyo;
 RX MEDLINE=96004459; PubMed=7551028;
 RA Takemitsu M., Matsumoto S., Ohara N., Kitaura H., Mizuno A.,
 RA Yamada T.;
 RA "Differential transcription of the MPB70 genes in two major groups of
 RT Mycobacterium bovis BCG substrains";
 RL Microbiology 141:1601-1607(1995).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett B., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RA "The complete genome sequence of Mycobacterium bovis";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 RN [8]
 RP SEQUENCE OF 49-193 FROM N.A.
 RC SPECIES=M.bovis;
 RX MEDLINE=8815076; PubMed=3278986;
 RA Radford A.J., Duffield B.J., Plackett P.;
 RA "Cloning of a species-specific antigen of Mycobacterium bovis";
 RL Infect. Immun. 56:921-925(1988).
 CC -1- SUBUNIT: GENERALLY FOUND AS A MONOMER; HOMODIMER IN CULTURE
 CC FLUIDS.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- MISCELLANEOUS: PRODUCED IN HIGH CONCENTRATION BY BCG TOKYO,
 CC MOREAU, RUSSIA AND SWEDEN (HIGH-PRODUCER SUBSTRAINS), WHEREAS IN
 CC BCG PASTEUR, COPENHAGEN AND TICE (LOW-PRODUCER SUBSTRAINS) IT IS
 CC DETECTED AT 1% (W/W) OR LESS OF THE CONCENTRATION OF BCG TOKYO.
 CC THE DIFFERENCE IN THE SECRETION BETWEEN BCG TOKYO AND PASTEUR WAS
 CC ATTRIBUTED TO DIFFERENTIAL TRANSCRIPTION EFFICIENCIES.
 CC -1- SIMILARITY: Contains 1 FAS1 domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; D37968; BAA07184.1; -
 CC EMBL; Z74024; CAA98373.1; -
 CC EMBL; AE007118; AAK47268.1; -
 CC EMBL; D38229; BAA07402.1; -
 CC EMBL; M33916; AAA25366.1; -
 CC EMBL; D38230; BAA07403.1; -
 CC EMBL; BX248344; CAD96887.1; -
 CC EMBL; A08199; CAA00760.1; -
 CC EMBL; M37840; AAA25355.1; -
 CC PIR; A37195; A37195.
 CC PIR; A48320; A48320.
 CC TIGR; M2943; -
 CC TubercuList; Rv2875; -
 CC InterPro; IPR000782; BIGH3_FAS1.
 CC

DR Pfam; PF02469; Fasciclin; 1.
 DR SMART; SM00554; FAS1; 1.
 DR PROSITE; PS0213; FAS1; 1.
 KW Antigen; Signal; Complete proteome.
 FT SIGNAL 1 30
 FT CHAIN 31 193 IMMUNOGENIC PROTEIN MPB70.
 FT DOMAIN 57 189 FAS1.
 FT CONFLICT 16 17 GL -> AV (IN REF. 6).
 FT CONFLICT 101 101 P -> R (IN REF. 8).
 SQ SEQUENCE 193 AA; 19072 MW; 228695731C3FFB00 CRC64;
 Query Match 100.0%; Score 95; DB 1; Length 193;
 Best Local Similarity 100.0%; Pred. No. 5.2e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GVSTANATVYMDSVLMPP 19
 DB 174 GVSTANATVYMDSVLMPP 192
 RESULT 2
 ID MP83 MYCTU STANDARD; PRT; 220 AA.
 AC Q10750; P71493;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 30-MAY-2000 (Rel. 35, Last sequence update)
 DE Cell surface lipoprotein MPB83 precursor (lipoprotein P23).
 GN MPB83 OR MPB83 OR RV2873 OR MTC2940 OR MTC274.04 OR MB2898.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_taxid=1773, 1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=96233689; PubMed=8633206;
 RA Hewinson R.G., Michell S., Russell W.P., McAdam R.A.,
 RA Jacobs W.R. Jr.;
 RA "Molecular characterization of MPB83: a seroreactive antigen of
 RT Mycobacterium tuberculosis with homology to MPB70";
 RL Scand. J. Immunol. 43:490-499(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence";
 RL Nature 393:537-544(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RL laboratory strains";
 RN J. Bacteriol. 184:5479-5490(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=BCG / Tokyo;

```

RX MEDLINE=96233688; PubMed=8633205;
RA Matsuo T., Matsuo H., Ohara N., Matsumoto S., Kitaura H., Mizuno A.,
RA Yamada T.;
RT "Cloning and sequencing of an MPB70 homologue corresponding to MPB83
RT from Mycobacterium bovis BCG.";
RL Scand. J. Immunol. 43:483-489(1996).
RN [5]
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF CVS-25.
RC SPECIES=M.bovis; STRAIN=BCG / Tokyo 172;
RX MEDLINE=97254460; PubMed=909870;
RA Vosloo W., Tippoo P., Hughes E.J., Harriman N., Emms M., Beatty D.W.,
RA Zappe H., Steyn L.M.;
RT "Characterisation of a lipoprotein in Mycobacterium bovis (BCG) with
RT sequence similarity to the secreted protein MPB70.";
RL Gene 188:123-128(1997).
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Dutchoy S., Grondin S., Lacroix C., Monsenpe C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.F., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor.
CC -!- MISCELLANEOUS: Highly immunogenic.
CC -!- SIMILARITY: Contains 1 FAS1 domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; X94597; CAA64290.1; -
CC EMBL; 274024; CAA98350.1; -
CC EMBL; AB007118; AAK47265.1; -
CC EMBL; D64165; BAA11027.1; -
CC EMBL; U28743; AAB03901.1; ALT_INIT.
CC EMBL; BX248344; CAD96585.1; -
CC PIR; D70923; D70923.
CC TIGR; MT2940; -
CC Tuberculin; Rv2873; -
CC InterPro; IPR000782; BIGH3_FAS1.
CC InterPro; IPR000437; Prok_lipoprot_S.
CC Pfam; PF02469; Fasciclin; 1.
CC SMART; SM00554; FAS1; 1.
CC PROSITE; PS50213; FAS1; 1.
CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Antigen; Lipoprotein; Membrane; Signal; Complete proteome; Palmitate.
FT SIGNAL 1 24
FT CHAIN 25 220 CELL SURFACE LIPOPROTEIN MPT83.
FT DOMAIN 83 215 FAS1.
FT LIPID 25 25 N-palmitoyl cysteine.
FT LIPID 25 25 S-diacylglycerol cysteine.
FT MUTAGEN C->S: LOSS OF ACYLATION.
FT SEQUENCE 220 AA; 22070 MW; 5CB999A4B51852A98 CRC64;
SQ
Query Match 91.6%; Score 87; DB 1; Length 220;
Best Local Similarity 89.5%; Pred. No. 1.4e-07;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GVSTANATVYMDIVLMP 19
Db 200 GVHTANATVYMDIVLMP 218
RESULT 3
YE83_SVNY3
ID_YE83_SVNY3 STANDARD; PRT; 180 AA.

```

```

AC P74615;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein sll1483 precursor.
GN SLL1483.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=9905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugitani M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- SIMILARITY: Contains 1 FAS1 domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; D90916; BAA18723.1; -
CC PIR; S76811; S76811.
CC InterPro; IPR000782; BIGH3_FAS1.
CC Pfam; PF02469; Fasciclin; 1.
CC SMART; SM00554; FAS1; 1.
CC PROSITE; PS50213; FAS1; 1.
CC PROSITE; PS00013; POTENTIAL.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 180 HYPOTHEICAL PROTEIN SLL1483.
FT DOMAIN 45 176 FAS1
SQ SEQUENCE 180 AA; 18471 MW; 1DE2D8A8AEE389 CRC64;
Query Match 51.6%; Score 49; DB 1; Length 180;
Best Local Similarity 38.9%; Pred. No. 0.4;
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
Qy 2 VSTANATVYMDIVLMP 19
Db 162 VDASNGVHVIDQVILPP 179
RESULT 4
BGH3_RABIT STANDARD; PRT; 683 AA.
AC Q35215;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transforming growth factor-beta induced protein IG-H3 precursor (Beta
DE IG-H3) (Kerato-epithelin) (RGD-containing collagen associated protein)
DE (RGD-CAP).
DE TGFBI.
GN Oryctolagus cuniculus (Rabbit).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Cornea;
RX MEDLINE=97267855; PubMed=9112985;
RA Rawe I.M., Zhan Q., Burrows R., Bennett K., Cintron C.;
RT "Beta-ig. Molecular cloning and in situ hybridization in corneal

```

RT Invest. Ophthalmol. Vis. Sci. 38:893-900(1997).
 RL FUNCTION: BINDS TO TYPE I, II, AND IV COLLAGENS. THIS ADHESION
 CC PROTEIN MAY PLAY AN IMPORTANT ROLE IN CELL-COLLAGEN INTERACTIONS.
 CC IN CARTILAGE, MAY BE INVOLVED IN ENDOCHONDRAL BONE FORMATION. MAY
 CC SUBCELLULAR LOCATION: Extracellular. May be associated both with
 CC microfibrils and with the cell surface (By similarity).
 CC TISSUE SPECIFICITY: LOCATED PRIMARILY IN THE EPITHELIUM OF NORMAL
 CC ADULT CORNEA, IN FETAL STROMAL CELLS, AND BOTH ENDOTHELIUM- AND
 CC STROMA-DERIVED CELLS IN HEALING CORNEAL WOUNDS. NOT EXPRESSED IN
 CC NORMAL ADULT ENDOTHELIUM AND STROMA.
 CC INDUCTION: By TGF-beta.
 CC SIMILARITY: Contains 4 FAS1 domains.
 CC CAUTION: Ref.1 sequence differs from that shown due to a
 CC frameshift in position 15.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: U66205; AAB07015.1; ALT FRAME.
 DR InterPro: IPR000782; BIGH3_FAS1.
 DR Pfam: PF02469; Fasciclin; 4.
 DR SMART: SMO0554; FAS1; 4.
 DR PROSITE: PS0213; FAS1; 4.
 DR Extracellular matrix; Signal; Repeat; Cell adhesion.
 FT SIGNAL 1 23 BY SIMILARITY.
 FT CHAIN 24 683 TRANSFORMING GROWTH FACTOR-BETA INDUCED
 FT DOMAIN 103 236 PROTEIN IG-H3.
 FT DOMAIN 240 371 FAS1 1.
 FT DOMAIN 375 498 FAS1 2.
 FT DOMAIN 502 632 FAS1 3.
 FT SITE 641 643 FAS1 4.
 FT SITE 641 643 CELL ATTACHMENT SITE (POTENTIAL).
 SQ SEQUENCE 683 AA; 74684 MW; 4548520497548CD6 CRC64;

Query Match 49.5%; Score 47; DB 1; Length 683;
 Best Local Similarity 50.0%; Pred. No. 3.6;
 Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
 QY 2 VSTANATVYMIDSVLMP 19
 Db 618 IMATNGVVAITSVLQPP 635

RESULT 5
 ID12_METH STANDARD; PRT; 349 AA.
 AC O26154;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Isopentenyl-diphosphate delta-isomerase (EC 5.3.3.2) (IPP isomerase)
 DE (Isopentenyl pyrophosphate isomerase).
 GN FNI OR MTH48.
 OS Methanobacterium thermoautotrophicum.
 CC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 CC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=167420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,

McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT deltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 CC FUNCTION: Catalyzes the 1,3-allylic rearrangement of the
 CC homoallylic substrate isopentenyl (IPP) to its allylic isomer,
 CC dimethylallyl diphosphate (DMAPP) (By similarity).
 CC CATALYTIC ACTIVITY: Isopentenyl diphosphate = dimethylallyl
 CC diphosphate.
 CC COFACTOR: FMN and NADPH (By similarity).
 CC SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC SIMILARITY: Belongs to the IPP isomerase type 2 family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: AE000797; AAB84555.1; -.
 DR PIR: H69162; H69162.
 DR HANAP: MF_00354; -; 1.
 DR InterPro: IPR003009; FMN enzyme.
 DR InterPro: IPR00262; FMN_hydxyc_dh.
 DR Pfam: PF01070; FMN_dh; 1.
 KW Isomerase; Isoprene biosynthesis; Flavoprotein; FMN; NADP;
 KW Complete proteome.
 SQ SEQUENCE 349 AA; 37142 MW; A73915E8E5B3A1E CRC64;
 Query Match 45.3%; Score 43; DB 1; Length 349;
 Best Local Similarity 44.4%; Pred. No. 8.7;
 Matches 8; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GVSANATVYMIDSVLMP 18
 Db 236 GIPTAASTVEVESVSP 253

RESULT 6
 SEFC_SALEN STANDARD; PRT; 814 AA.
 AC P33388;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Outer membrane usher protein sefc precursor.
 GN SEFC.
 OS Salmonella enteritidis.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=592;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=27655-3B;
 RX MEDLINE=93239677; PubMed=8097515;
 RA Clouthier S.C., Mueller K.-H., Doran J.L., Collinson S.K., Kay W.W.;
 RT "Characterization of three fimbrial genes, sefABC, of Salmonella
 RT enteritidis";
 RL J. Bacteriol. 175:2523-2533(1993).
 CC FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF THE SEFA
 CC FIMBRIAL SUBUNIT.
 CC SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
 CC (By similarity).
 CC SIMILARITY: Belongs to the fimbrial export usher family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [5]
RN TISSUE SPECIFICITY.
RX MEDLINE=94357992; PubMed=8077289;
RA Escaribano J., Hernandez N., Ghosh S., Crabb J., Coca-Prados M.;
RA "cDNA from human ocular ciliary epithelium homologous to beta ig-h3 is
RT preferentially expressed as an extracellular protein in the corneal
RT epithelium";
RL J. Cell. Physiol. 160:511-521 (1994).
RN [6]
RN REVIEW ON VARIANTS CORNEAL DYSTROPHIES.
RX MEDLINE=21392849; PubMed=11501939;
RA Fujiki K., Nakayasu K., Kanai A.;
RT "Corneal dystrophies in Japan";
RL J. Hum. Genet. 46:431-435 (2001).
RN [7]
RN VARIANTS CORNEAL DYSTROPHIES CYS-124; HIS-124; GLN-555 AND TRP-555.
RX MEDLINE=97207642; PubMed=9054935;
RA Munier F.L., Korvatska E., Djemai A., le Paslier D., Zografos L.,
RA Pescia G., Schorderet D.F.;
RT "Kerato-epithelin mutations in four 5q31-linked corneal dystrophies";
RL Nat. Genet. 15:247-251 (1997).
RN [8]
RN VARIANTS CORNEAL DYSTROPHIES CYS-124; HIS-124; GLN-555 AND TRP-555.
RX MEDLINE=98130534; PubMed=9463327;
RA Korvatska E., Munier F.L., Djemai A., Wang M.X., Frueh B.,
RA Chlou A.G.-Y., Uffer S., Ballestazzi E., Braunstein R.E.,
RA Forster R.K., Culbertson W.W., Bonan H., Zografos L., Schorderet D.F.;
RT "Mutation hot spots in 5q31-linked corneal dystrophies";
RL Am. J. Hum. Genet. 62:320-324 (1998).
RN [9]
RN VARIANT CDL3A THR-501.
RX MEDLINE=98163459; PubMed=9497262;
RA Yamamoto S., Okada M., Tsujikawa M., Shimomura Y., Nishida K.,
RA Inoue Y., Watanabe H., Maeda N., Kurahashi H., Kinoshita S.,
RA Nakamura Y., Tano Y.;
RT "A kerato-epithelin (beta-ig-h3) mutation in lattice corneal dystrophy
RL type IIIA";
RL Am. J. Hum. Genet. 62:719-722 (1998).
RN [10]
RN VARIANT CDRB LEU-124.
RX MEDLINE=98451378; PubMed=9780098;
RA Okada M., Yamamoto S., Tsujikawa M., Watanabe H., Inoue Y., Maeda N.,
RA Shimomura Y., Nishida K., Quantock A.J., Kinoshita S., Tano Y.;
RT "Two distinct kerato-epithelin mutations in Reis-Buecklers corneal
RL dystrophy";
RL Am. J. Ophthalmol. 126:535-542 (1998).
RN [11]
RN VARIANT CDL1 ARG-527.
RX MEDLINE=99013426; PubMed=9799082;
RA Fujiki K., Hotta Y., Nakayasu K., Yokoyama T., Takano T.,
RA Yamaguchi T., Kanai A.;
RT "A new L527R mutation of the betaIGH3 gene in patients with lattice
RL corneal dystrophy with deep stromal opacities";
RL Hum. Genet. 103:286-289 (1998).
RN [12]
RN VARIANT CDRB PHE-540 DEL.
RA Rozzo C., Fossarello M., Galleri G., Sole G., Serru A., Orzalesi N.,
RA Serra A., Pirastu M.;
RT "A common beta ig-h3 gene mutation (delta F540) in a large cohort of
RL Sardinian Reis-Buecklers' corneal dystrophy patients";
RL Hum. Mutat. 12:215-216 (1998).
RN [13]
RN VARIANTS CORNEAL DYSTROPHIES HIS-124; SER-124 AND TRP-555.
RX MEDLINE=99355712; PubMed=10425035;
RA Stewart H.S., Ridgway A.E., Dixon M.J., Bonshek R.E., Parveen R.,
RA Black G.C.;
RT "Heterogeneity in granular corneal dystrophy: identification of three
RT Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RT Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RT Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RN "Generation and initial analysis of more than 15,000 full-length
RN human and mouse cDNA sequences";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [5]
RN TISSUE SPECIFICITY.
RX MEDLINE=94357992; PubMed=8077289;
RA Escaribano J., Hernandez N., Ghosh S., Crabb J., Coca-Prados M.;
RA "cDNA from human ocular ciliary epithelium homologous to beta ig-h3 is
RT preferentially expressed as an extracellular protein in the corneal
RT epithelium";
RL J. Cell. Physiol. 160:511-521 (1994).
RN [6]
RN REVIEW ON VARIANTS CORNEAL DYSTROPHIES.
RX MEDLINE=21392849; PubMed=11501939;
RA Fujiki K., Nakayasu K., Kanai A.;
RT "Corneal dystrophies in Japan";
RL J. Hum. Genet. 46:431-435 (2001).
RN [7]
RN VARIANTS CORNEAL DYSTROPHIES CYS-124; HIS-124; GLN-555 AND TRP-555.
RX MEDLINE=97207642; PubMed=9054935;
RA Munier F.L., Korvatska E., Djemai A., le Paslier D., Zografos L.,
RA Pescia G., Schorderet D.F.;
RT "Kerato-epithelin mutations in four 5q31-linked corneal dystrophies";
RL Nat. Genet. 15:247-251 (1997).
RN [8]
RN VARIANTS CORNEAL DYSTROPHIES CYS-124; HIS-124; GLN-555 AND TRP-555.
RX MEDLINE=98130534; PubMed=9463327;
RA Korvatska E., Munier F.L., Djemai A., Wang M.X., Frueh B.,
RA Chlou A.G.-Y., Uffer S., Ballestazzi E., Braunstein R.E.,
RA Forster R.K., Culbertson W.W., Bonan H., Zografos L., Schorderet D.F.;
RT "Mutation hot spots in 5q31-linked corneal dystrophies";
RL Am. J. Hum. Genet. 62:320-324 (1998).
RN [9]
RN VARIANT CDL3A THR-501.
RX MEDLINE=98163459; PubMed=9497262;
RA Yamamoto S., Okada M., Tsujikawa M., Shimomura Y., Nishida K.,
RA Inoue Y., Watanabe H., Maeda N., Kurahashi H., Kinoshita S.,
RA Nakamura Y., Tano Y.;
RT "A kerato-epithelin (beta-ig-h3) mutation in lattice corneal dystrophy
RL type IIIA";
RL Am. J. Hum. Genet. 62:719-722 (1998).
RN [10]
RN VARIANT CDRB LEU-124.
RX MEDLINE=98451378; PubMed=9780098;
RA Okada M., Yamamoto S., Tsujikawa M., Watanabe H., Inoue Y., Maeda N.,
RA Shimomura Y., Nishida K., Quantock A.J., Kinoshita S., Tano Y.;
RT "Two distinct kerato-epithelin mutations in Reis-Buecklers corneal
RL dystrophy";
RL Am. J. Ophthalmol. 126:535-542 (1998).
RN [11]
RN VARIANT CDL1 ARG-527.
RX MEDLINE=99013426; PubMed=9799082;
RA Fujiki K., Hotta Y., Nakayasu K., Yokoyama T., Takano T.,
RA Yamaguchi T., Kanai A.;
RT "A new L527R mutation of the betaIGH3 gene in patients with lattice
RL corneal dystrophy with deep stromal opacities";
RL Hum. Genet. 103:286-289 (1998).
RN [12]
RN VARIANT CDRB PHE-540 DEL.
RA Rozzo C., Fossarello M., Galleri G., Sole G., Serru A., Orzalesi N.,
RA Serra A., Pirastu M.;
RT "A common beta ig-h3 gene mutation (delta F540) in a large cohort of
RL Sardinian Reis-Buecklers' corneal dystrophy patients";
RL Hum. Mutat. 12:215-216 (1998).
RN [13]
RN VARIANTS CORNEAL DYSTROPHIES HIS-124; SER-124 AND TRP-555.
RX MEDLINE=99355712; PubMed=10425035;
RA Stewart H.S., Ridgway A.E., Dixon M.J., Bonshek R.E., Parveen R.,
RA Black G.C.;
RT "Heterogeneity in granular corneal dystrophy: identification of three
RT caustic mutations in the TGFB1 (BIGH3) gene-lessons for corneal
RT amyloidogenesis";
RL Hum. Mutat. 14:126-132 (1999).
RN [14]
RN VARIANTS CDL HIS-622 AND ARG-626.
RX MEDLINE=99258949; PubMed=10328397;
RA Stewart H.S., Black G.C., Donnai D., Bonshek R.E., McCarthy J.,
RA Morgan S., Dixon M.J., Ridgway A.A.;
RT "A mutation within exon 14 of the TGFB1 (BIGH3) gene on chromosome
RT 5q31 causes an asymmetric, late-onset form of lattice corneal
RT dystrophy";
RL Ophthalmology 106:964-970 (1999).
RN [15]
RN VARIANTS CORNEAL DYSTROPHIES CYS-124; HIS-124; LEU-124; THR-501;
RX ARG-527; SER-544; GLN-555 AND TRP-555.
RX MEDLINE=20480024; PubMed=11024425;
RA Mashima Y., Yamamoto S., Inoue Y., Yamada M., Konishi M., Watanabe H.,
RA Maeda N., Shimomura Y., Kinoshita S.;
RT "Association of autosomal dominantly inherited corneal dystrophies
RT with BIGH3 gene mutations in Japan";
RL Am. J. Ophthalmol. 130:516-517 (2000).
RN [16]
RN VARIANTS CORNEAL DYSTROPHIES LEU-124 AND 125-THR-GLU-126 DEL.
RX MEDLINE=20325588; PubMed=10865320;
RA Dighiero P., Drunat S., D'Hermies F., Renard G., Delpech M.,
RA Vallex S.;
RT "A novel variant of granular corneal dystrophy caused by association
RT of 2 mutations in the TGFB1 gene-R124L and delta125-delta126";
RL Arch. Ophthalmol. 118:814-819 (2000).
RN [17]
RN VARIANT CDL1 PRO-518.
RX MEDLINE=20298440; PubMed=10837380;
RA Hirano K., Hotta Y., Fujiki K., Kanai A.;
RT "Corneal amyloidosis caused by Leu518Pro mutation of betaig-h3 gene";
RL Br. J. Ophthalmol. 84:583-585 (2000).
RN [18]
RN VARIANT CDL1 ARG-527.
RX MEDLINE=21305893; PubMed=11413411;
RA Hirano K., Hotta Y., Nakamura M., Fujiki K., Kanai A., Yamamoto N.;
RT "Late-onset form of lattice corneal dystrophy caused by Leu527Arg
RL mutation of the TGFB1 gene";
RL Cornea 20:525-529 (2001).
RN [19]
RN VARIANTS CORNEAL DYSTROPHIES CYS-124; HIS-124; LEU-124;
RX 125-THR-GLU-126 DEL; THR-546; GLN-555; TRP-555 AND ARG-626.
RX MEDLINE=21196035; PubMed=11297504;
RA Dighiero P., Niel F., Ellies P., D'Hermies F., Savoldelli M.,
RA Renard G., Delpech M., Vallex S.;
RT "Histologic phenotype-genotype correlations of corneal dystrophies
RT associated with eight distinct mutations in the TGFB1 gene";
RL Ophthalmology 108:818-823 (2001).
RN [20]
RN FUNCTION: Binds to type I, II, and IV collagens. This adhesion
RN protein may play an important role in cell-collagen interactions.
RN In cartilage, may be involved in endochondral bone formation.
RN SUBCELLULAR LOCATION: Extracellular. May be associated both with
RN microfibrils and with the cell surface.
RN TISSUE SPECIFICITY: Highly expressed in the corneal epithelium.
RN INDUCTION: By TGF-beta.
RN DISEASE: Defects in TGFB1 are the cause of corneal dystrophy
RN Groenouw type I (CDG1) [MIM:121900]; also known as corneal
RN dystrophy granular type. Inheritance is autosomal dominant.
RN Corneal dystrophies show progressive opacification of the cornea
RN leading to severe visual handicap.
RN DISEASE: Defects in TGFB1 are the cause of Reis-Buecklers corneal
RN dystrophy (CDRB) [MIM:121900].
RN DISEASE: Defects in TGFB1 are the cause of corneal dystrophy
RN lattice type I (CDL1) [MIM:122200]. Inheritance is autosomal
RN dominant.
RN DISEASE: Defects in TGFB1 are the cause of lattice corneal
RN dystrophy type IIIA (CDL3A) [MIM:204870]. CDL3A clinically
RN resembles to lattice corneal dystrophy type III, but differs in
RN that its age of onset is 70 to 90 years. It has an autosomal
RN dominant inheritance pattern.

CC -I- DISEASE: Defects in TGFBI are the cause of Avellino corneal
 CC dystrophy (ACD) [MIM:607541]. ACD could be considered a variant of

Query Match 43.2%; Score 41; DB 1; Length 683;
 Best Local Similarity 38.9%; Pred. No. 39;
 Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 VSTANATVYMIDSVLMPP 19
 DB 618 IMATNGVWVITNVLP 635

RESULT 9
 GAK_RAT
 ID GAK_RAT STANDARD; PRT; 1305 AA.
 AC P97874;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cyclin G-associated kinase (EC 2.7.1.1-).
 GN GAK.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=97165969; PubMed=9013862;
 RA Kanaoka Y., Kimura S.H., Okazaki I., Ikeda M., Nojima H.;
 RT "GAK: a cyclin G associated kinase contains a tensin/auxilin-like
 domain.";
 RL FEBS Lett. 402:73-80(1997).
 CC -I- FUNCTION: Associates with cyclin G and CDKs. Seems to act as an
 CC auxilin homolog that is involved in the uncoating of clathrin-
 CC coated vesicles by Hsc70 in non-neuronal cells. Expression
 CC oscillates slightly during the cell cycle, peaking at G1 (By
 CC similarity).
 CC -I- SUBCELLULAR LOCATION: Localizes to the perinuclear area and to the
 CC trans-Golgi network. Also seen on the plasma membrane, probably at
 CC focal adhesions (By similarity).
 CC -I- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
 CC -I- SIMILARITY: Contains 1 tensin domain.
 CC -I- SIMILARITY: Contains 1 J domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D38560; BAA18911.1; -;
 CC PIR; T31096; T31096.
 CC InterPro; IPR008973; C2_CaLb.
 CC InterPro; IPR001623; DnaJ_N.
 CC InterPro; IPR000719; Prot_kinase.
 CC InterPro; IPR008271; Ser_thr_pkin_AS.
 CC Pfam; PF00226; DnaJ; 1.
 CC Pfam; PF00069; pkinase; 1.
 CC ProDom; PD000001; Prot_kinase; 1.
 CC SMART; SM00271; DnaJ; 1.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 CC PROSITE; PS00636; DnaJ_1; FALSE_NEG.
 CC PROSITE; PS50076; DnaJ_2; 1.
 CC Transferrase; Serine/threonine-protein kinase; ATP-binding;
 CC Nuclear protein; Endoplasmic reticulum; Cell cycle.
 FT DOMAIN 40 315 PROTEIN_KINASE.
 FT ACT_SITE 173 173 BY SIMILARITY.
 FT TENSIN 405 689

FT DOMAIN 1241 1305 J-DOMAIN.
 SQ SEQUENCE 1305 AA; 143702 MW; 6D36BD38011C44EE CRC64;

Query Match 43.2%; Score 41; DB 1; Length 1305;
 Best Local Similarity 52.9%; Pred. No. 75;
 Matches 9; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 3 STANATVYMIDSVLMPP 19
 DB 528 STAEAAVYMFMSMKRCP 544

RESULT 10
 GAK_HUMAN
 ID GAK_HUMAN STANDARD; PRT; 1311 AA.
 AC O14976; Q9BVY6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cyclin G-associated kinase (EC 2.7.1.1-).
 GN GAK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=Fibroblast;
 RX MEDLINE=97446136; PubMed=9299234;
 RA Kimura S.H., Tsuruga H., Yabuta N., Endo Y., Nojima H.;
 RT "Structure, expression, and chromosomal localization of human GAK.";
 RL Genomics 44:179-187(1997).
 RN [2]
 RP SEQUENCE OF 981-1311 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=92388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Stapleton L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Richards S., Worley K.C., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Pahey J., Hilton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP FUNCTION, AND SUBCELLULAR LOCATION.
 RX MEDLINE=20092914; PubMed=10625686;
 RA Greener T., Zhao X., Nojima H., Eisenberg E., Greene L.E.;
 RT "Role of cyclin G-associated kinase in uncoating clathrin-coated
 RT vesicles from non-neuronal cells.";
 RL J. Biol. Chem. 275:1365-1370(2000).
 CC -I- FUNCTION: Associates with cyclin G and CDKs. Seems to act as an
 CC auxilin homolog that is involved in the uncoating of clathrin-
 CC coated vesicles by Hsc70 in non-neuronal cells. Expression
 CC oscillates slightly during the cell cycle, peaking at G1.
 CC -I- SUBCELLULAR LOCATION: Localizes to the perinuclear area and to the
 CC trans-Golgi network. Also seen on the plasma membrane, probably at
 CC focal adhesions.
 CC -I- TISSUE SPECIFICITY: Ubiquitous. Highest in testis.
 CC -I- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
 CC -I- SIMILARITY: Contains 1 tensin domain.


```

CC CC      -!- SIMILARITY: Contains 1 J domain.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
CC CC      EMBL: D88435; BAA22623.1; -
CC CC      EMBL: BC008016; AAH00816.1; -
CC CC      EMBL: BC008668; AAH08668.1; -
CC CC      Genew: HGNC:4113; GAK.
CC CC      MIN: 602052; -
CC CC      GO: GO:0004674; F:protein serine/threonine kinase activity; TAS.
CC CC      GO: GO:0000074; P:regulation of cell cycle; TAS.
CC CC      InterPro: IPR008973; C2 CALB.
CC CC      InterPro: IPR001623; DnaJ_N.
CC CC      InterPro: IPR00719; Prot_kinase.
CC CC      InterPro: IPR008271; Ser_Chkr_pkin_AS.
CC CC      Pfam: PF00226; DnaJ; 1.
CC CC      Pfam: PF00069; pkinase; 1.
CC CC      SMART: PD000001; Prot_kinase; 1.
CC CC      SMART: SC00271; DnaJ; 1.
CC CC      PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
CC CC      PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
CC CC      PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
CC CC      PROSITE: PS00636; DnaJ_1; FALSE_NEG.
CC CC      PROSITE: PS00076; DnaJ_2; 1.
CC CC      PROSITE: PS00076; DnaJ_2; 1.
CC CC      Transferase; Serine/threonine-protein kinase; ATP-binding;
CC CC      Nuclear protein; Endoplasmic reticulum; Cell cycle.
CC CC      DOMAIN 40 314
CC CC      ACT SITE 173 373 BY SIMILARITY.
CC CC      DOMAIN 347 350 POLY-PRO.
CC CC      TENSIN.
CC CC      DOMAIN 1247 1311 J-DOMAIN.
CC CC      CONFLICT 1113 1113 P -> A (IN REF. 1).
CC CC      SQ SEQUENCE 1311 AA; 143190 MW; 0ACE45DF57A5F981 CRC64;

Query Match 43.28; Score 41; DB 1; Length 1311;
Best Local Similarity 52.99; Pred. No. 76;
Matches 9; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Oy 3 STANATYVIMDSVLMP 19
Db 530 STREAAYVFMKRCPP 546

RESULT 11
ID DNMI_MOUSE STANDARD; PRT; 1620 AA.
AC P13864; P97413; Q9CSG6; Q9QXX6;
DT 01-JAN-1990 (Rel. 13, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA (cytosine-5'-methyltransferase 1 (EC 2.1.1.37) (Dnmt1) (DNA
DE methyltransferase MnuI) (DNA M5ase MnuI) (MCMT) (M.MnuI) (Met-1).
GN DNMT1 OR DNMT OR UIM OR MET1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=89094873; PubMed=3210246;
RA Bestor T.H., Laudano A., Mattaliano R., Ingram V.;
RT "Cloning and sequencing of a cDNA encoding DNA methyltransferase of
RT mouse cells. The carboxyl-terminal domain of the mammalian enzymes is
RT related to bacterial restriction methyltransferases."
RL J. Mol. Biol. 203:971-983(1988).
RN [2]
RP REVISIONS TO N-TERMINUS.

```

```

RC TISSUE=Embryo;
RX MEDLINE=97094871; PubMed=8940105;
RA Yoder J.A., Yen R.-W.C., Vertino P.M., Bestor T.H., Bayliss S.B.;
RT "New 5' regions of the murine and human genes for DNA (cytosine-5)-
RT methyltransferase."
RL J. Biol. Chem. 271:31092-31097(1996).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX STRAIN=C57BL/6; TISSUE=Skeletal muscle;
RX MEDLINE=20515133; PubMed=11063128;
RA Aguirre-Arteta A.M., Grunewald I., Cardoso M.C., Leonhardt H.;
RT "Expression of an alternative Dnmt1 isoform during muscle
RT differentiation."
RL Cell Growth Differ. 11:551-559(2000).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX STRAIN=C57BL/6;
RX MEDLINE=20181859; PubMed=10715201;
RA Margot J.B., Aguirre-Arteta A.M., Di Giacco B.V., Pradhan S.,
RA Roberts R.J., Cardoso M.C., Leonhardt H.;
RT "Structure and function of the mouse DNA methyltransferase gene: Dnmt1
RT shows a tripartite structure."
RL J. Mol. Biol. 297:293-300(2000).
RN [5]
RP SEQUENCE OF 1-27 AND 119-1619 FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=98119799; PubMed=9449671;
RA Mertineit C., Yoder J.A., Taketo T., Laird D.W., Trasler J.M.,
RA Bestor T.H.;
RT "Sex-specific exons control DNA methyltransferase in mammalian germ
RT cells."
RL Development 125:889-897(1998).
RN [6]
RP SEQUENCE OF 1-144 FROM N.A. (ISOFORMS 1 AND 2), AND SEQUENCE OF 3-6.
RX STRAIN=129/SV, and BALB/c; TISSUE=Embryonic stem cells;
RX MEDLINE=99047652; PubMed=9830015;
RA Gaudet F., Talbot D., Leonhardt H., Jaenisch R.;
RT "A short DNA methyltransferase isoform restores methylation in vivo."
RL J. Biol. Chem. 273:32725-32729(1998).
RN [7]
RP SEQUENCE OF 1-119 FROM N.A. (ISOFORM 1).
RX STRAIN=129/SV; TISSUE=Embryonic stem cells, and Kidney;
RX MEDLINE=97075093; PubMed=8917520;
RA Tucker K.L., Talbot D., Lee M.A., Leonhardt H., Jaenisch R.;
RT "Complementation of methylation deficiency in embryonic stem cells by
RT a DNA methyltransferase minigene."
RL Proc. Natl. Acad. Sci. U.S.A. 93:12920-12925(1996).
RN [8]
RP SEQUENCE OF 1-272 FROM N.A. (ISOFORM 1).
RX STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischman W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guenichon S., Hill D., Hofmann J., Mashima J., Mazzarelli K., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli K., Kamiya M., Lee N.H.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohteki S.,
RA Havaashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
RN [9]
RP PHOSPHORYLATION OF SER-515, AND MASS SPECTROMETRY.
RP TISSUE=Erythrocyte;

```

RX MEDLINE=97362284; PubMed=9211941;
 RA Glickman J.F., Pavlovich J.G., Reich N.O.;
 RT "Peptide mapping of the murine DNA methyltransferase reveals a major
 RT phosphorylation site and the start of translation.";
 J. Biol. Chem. 272:17851-17857(1997).
 [10]
 RN INTERACTION WITH HDAC1.
 RP MEDLINE=40082816; PubMed=10615135;
 RA Fuks F., Burgers W.A., Brhm A., Hughes-Davies L., Kouzarides T.;
 RT "DNA methyltransferase Dnmt1 associates with histone deacetylase
 RT activity.";
 Nat. Genet. 24:88-91(2000).
 [11]
 RN INTERACTION WITH HDAC2 AND DNAP1.
 RP MEDLINE=20347709; PubMed=10888872;
 RA Rountree M.R., Bachman K.E., Baylin S.B.;
 RT "DNMT1 binds HDAC2 and a new co-repressor, DNAP1, to form a complex at
 RT replication foci.";
 Nat. Genet. 25:269-277(2000).
 [12]
 RN FUNCTION, AND SUBCELLULAR LOCATION.
 RP MEDLINE=21185930; PubMed=11290321;
 RA Howell C.Y., Bestor T.H., Ding F., Latham K.E., Mertins C.,
 RA Trasler J.M., Chaillet J.R.;
 RT "Genomic imprinting disrupted by a maternal effect mutation in the
 RT Dnmt1 gene.";
 Cell 104:829-838(2001).
 [13]
 RN ALLOSTERIC ACTIVATION.
 RP MEDLINE=2129215; PubMed=11399088;
 RA Fatem M., Hermann A., Pradhan S., Jeltsch A.;
 RT "The activity of the murine DNA methyltransferase Dnmt1 is controlled
 RT by interaction of the catalytic domain with the N-terminal part of
 RT the enzyme leading to an allosteric activation of the enzyme after
 RT binding to methylated DNA.";
 J. Mol. Biol. 309:1189-1199(2001).
 CC -|- FUNCTION: Methylates CpG residues. Preferentially methylates
 CC hemimethylated DNA. It is responsible for maintaining methylation
 CC patterns established in development. Isoform 2, in oocytes, may
 CC provide maintenance methyltransferase activity specifically at
 CC imprinted loci during the fourth embryonic S phase. Mediates
 CC transcriptional repression by direct binding to HDAC2.
 CC -|- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA = S-adenosyl-L-
 CC homocysteine + DNA containing 5-methylcytosine.
 CC -|- ENZYME REGULATION: Allosterically regulated. The binding of 5-
 CC methylcytosine-containing DNA to the N-terminal parts of Dnmt1
 CC causes an allosteric activation of the catalytic domain by a
 CC direct interaction of its Zn-binding domain with the catalytic
 CC domain.
 CC -|- SUBUNIT: Interacts with HDAC1 and with PCNA. Forms a complex with
 CC DNAP1 and HDAC2, with direct interaction.
 CC -|- SUBCELLULAR LOCATION: Nuclear; it is nucleoplasmic through most of
 CC the cell cycle and associates with replication foci during S-
 CC phase. In germ cells: spermatogonia, preleptotene and leptotene
 CC spermatocytes all express high levels of nuclear protein, while
 CC the protein is not detected in pachytene spermatocytes, despite
 CC the fact they expressed high levels of mRNA. In females, the
 CC protein is not detected in non-growing oocytes, in contrast to the
 CC growing oocytes. During the growing, the protein is no longer
 CC detectable in nuclei but accumulates to very high levels first
 CC throughout the cytoplasm. At the time of ovulation, all the
 CC protein is cytoplasmic and is actively associated with the oocyte
 CC cortex. After fecundation, in the preimplantation embryo, the
 CC protein remains cytoplasmic and after implantation, it is
 CC exclusively nuclear in all tissue types. Isoform 2 is sequestered
 CC in the cytoplasm of maturing oocytes and of preimplantation
 CC embryos, except for the 8-cell stage, while isoform 1 is
 CC exclusively nuclear.
 CC -|- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=Long;
 CC IsoId=P13864-1; Sequence=Displayed;
 CC Name=2; Synonyms=Short;

CC IsoId=P13864-2; Sequence=VSP 005619;
 CC -|- TISSUE SPECIFICITY: Isoform 1 is expressed in embryonic stem cells
 CC and in somatic tissues. Isoform 2 is expressed in oocytes,
 CC preimplantation embryos, testis and in skeletal muscle during
 CC myogenesis.
 CC -|- DEVELOPMENTAL STAGE: In germ cells, it is present at high levels
 CC in spermatogonia and spermatocytes until the pachytene stage,
 CC where it fails to undetectable levels. The transient drop at the
 CC pachytene stage coincides with the disappearance of the 5.2 kb
 CC mRNA and the accumulation of a larger 6.0 kb mRNA. Oocytes
 CC accumulate very large amounts of Dnmt1 protein during the growth
 CC phase.
 CC -|- MISCELLANEOUS: There are three 5' exons, one specific to the
 CC oocyte (1c), one specific to the pachytene spermatocyte and also
 CC found in skeletal muscle (1b) and one found in somatic cells (1a).
 CC Three different mRNAs can be produced which give rise to two
 CC different translation products: isoform 1 (mRNAs-1a) and isoform 2
 CC (mRNA-1b or -1c).
 CC -|- SIMILARITY: Belongs to the C5-methyltransferase family.
 CC -|- SIMILARITY: Contains 2 BAH domains.
 CC -|- SIMILARITY: Contains 1 CXXC-type zinc finger.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcement/>
 CC or send an email to license@ebi.ac.uk).
 CC EMBL; X14805; CAA32910.1; --
 CC EMBL; AF175432; AAF37695.1; --
 DR Query Match 42.6%; Score 40.5; DB 1; Length 1620;
 DR Best Local Similarity 52.9%; Pred. No. 1.2e+02;
 DR Matches 9; Conservative 2; Mismatches 5; Indels 1; Gaps 1;
 QY 4 TANATVYMI-DSVLMP 19
 DB 931 TKGWVRLGDSVILPP 947
 RESULT 12
 DNMT1 RAT STANDARD; PRT: 1622 AA.
 AC Q92330; P70487; Q9R252; Q9WTX3; Q9WU57;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA (cytosine-5)-methyltransferase 1 (EC 2.1.1.37) (Dnmt1) (DNA
 DE methyltransferase 1) (DNA MTase RnoIP) (MCMT) (M.RnoIP).
 GN DNMT1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Placenta;
 RX MEDLINE=9907263; PubMed=9878564;
 RA Kimura H., Takeda T., Tanaka S., Ogawa T., Shiota K.;
 RT "Expression of rat DNA (cytosine-5) methyltransferase (DNA MTase) in
 RT rodent trophoblast giant cells: molecular cloning and characterization
 RT of rat DNA MTase.";
 RL Biochem. Biophys. Res. Commun. 253:495-501(1998).
 [2]
 RP SEQUENCE OF 1-144 FROM N.A. (ISOFORMS 1 AND 9).
 RC TISSUE=Brain;
 RA Deng J., Szyf M.;
 RT "Multiple N-terminal isoforms of DNA (cytosine-5)-methyltransferase
 RT in vivo.";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 [3]

SEQUENCE OF 17-356 FROM N.A., AND IN VITRO BINDING TO ANNEXIN V.
 STRAIN=Wistar; TISSUE=Brain;
 MEDLINE=96301899; PubMed=8667030;
 Ohsawa K., Imai Y., Ito D., Kohaka S.;
 "Molecular cloning and characterization of annexin V-binding proteins
 with highly hydrophilic peptide structure.";
 J. Neurochem. 67:89-97(1996).
 [4]
 SEQUENCE OF 1169-1517 FROM N.A. (ISOFORMS 1; 2; 3; 4; 5; 6; 7 AND 8).
 MEDLINE=98389705; PubMed=9722504;
 Deng J., Szef M.;
 "Multiple isoforms of DNA methyltransferase are encoded by the
 vertebrate cytosine DNA methyltransferase gene.";
 J. Biol. Chem. 273:22869-22872(1998).
 CC -!- FUNCTION: Methylates CpG residues. Preferentially methylates
 hemimethylated DNA. It is responsible for maintaining methylation
 patterns established in development (By similarity). Mediates
 transcriptional repression by direct binding to HDAC2 (By
 similarity).
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA = S-adenosyl-L-
 homocysteine + DNA containing 5-methylcytosine.
 CC -!- SUBUNIT: Binds annexin V (potential).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=9;
 Comment=Additional isoforms seem to exist;
 Name=1;
 IsoId=Q92330-1; Sequence=Displayed;
 Name=2; Synonyms=SR1;
 IsoId=Q92330-2; Sequence=VSP_005627;
 Name=3; Synonyms=SR2;
 IsoId=Q92330-3; Sequence=VSP_005625;
 Name=4; Synonyms=SR3;
 IsoId=Q92330-4; Sequence=VSP_005624;
 Name=5; Synonyms=SR4;
 IsoId=Q92330-5; Sequence=VSP_005621;
 Name=6; Synonyms=SR5;
 IsoId=Q92330-6; Sequence=VSP_005622;
 Name=7; Synonyms=SR6;
 IsoId=Q92330-7; Sequence=VSP_005626;
 Name=8; Synonyms=SR7;
 IsoId=Q92330-8; Sequence=VSP_005623;
 Name=9; Synonyms=short;
 IsoId=Q92330-9; Sequence=VSP_005620;
 CC -!- TISSUE SPECIFICITY: Isoforms 0 and 8 are highly expressed in
 placenta, brain, lung, spleen, kidney, heart, and at much lower
 levels in liver. Isoform 1 is expressed in cerebellum, isoform 2
 in muscle and testis, isoform 3 in lung, isoform 4 in spleen and
 brain, and isoform 5 in brain.
 CC -!- SIMILARITY: Belongs to the C5-methyltransferase family.
 CC -!- SIMILARITY: Contains 2 BAH domains.
 CC -!- SIMILARITY: Contains 1 CXXC-type zinc finger.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; AB012214; BAA37118.1; -;
 EMBL; AF116344; AAD32541.1; -;
 EMBL; AF116345; AAD32542.1; -;
 EMBL; D64060; BAA20854.1; -;
 EMBL; AF083043; AAD28102.1; -;
 EMBL; AF083038; AAD28102.1; JOINED.
 EMBL; AF083039; AAD28102.1; JOINED.
 EMBL; AF083040; AAD28102.1; JOINED.
 EMBL; AF083041; AAD28102.1; JOINED.
 EMBL; AF083042; AAD28102.1; JOINED.
 EMBL; AF083043; AAD28102.1; JOINED.
 F1; J03078; J03078.
 REBASE; 3019; M.RncIP.

DR InterPro; IPR001025; BAH.
 DR InterPro; IPR001525; C5 DNA meth.
 DR InterPro; IPR002857; ZnF_CXXC.
 DR Pfam; PF01426; BAH; 2
 DR Pfam; PF00145; DNA_methylase; 1.
 DR Pfam; PF02008; zf-CXXC; 1.
 DR PRINTS; PR00105; CSMETFRASE.
 DR SMART; SMO0439; BAH; 2.
 DR TIGRFAMs; TIGR00675; dcm; 1.
 DR PROSITE; PS00094; C5_MTASE_1; 1.
 DR PROSITE; PS00095; C5_MTASE_2; 1.
 KW Transferase; Methyltransferase; Transcription regulation; Repressor;
 KW DNA-binding; Zinc-finger; Zinc; Metal-binding; Nuclear protein;
 KW Repeat; Phosphorylation; Alternative splicing.
 FT DOMAIN 173 200 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 327 556 DNA REPLICATION FOCI-TARGETING SEQUENCE.
 FT ZN FING 649 695 CXXC-TYPE.
 FT DOMAIN 759 884 BAH 1.
 FT DOMAIN 935 1105 BAH 2.
 FT DOMAIN 1114 1126 6.5 X 2 AA TANDEM REPEATS OF K-G.
 FT DOMAIN 1144 1622 CATALYTIC.
 FT DOMAIN 149 152 POLY-SER.
 FT DOMAIN 269 274 POLY-ASP.
 FT DOMAIN 726 729 POLY-LYS.
 FT ACT_SITE 1231 1231 BY SIMILARITY.
 FT MOD_RES 515 515 PHOSPHORYLATION (BY SIMILARITY).
 FT VARSPLIC 1 118 Missing (in isoform 9).
 FT VARSPLIC 1218 1430 /FTid=VSP_005620.
 FT VARSPLIC 1202 1410 QKGVEMLCGGPPCGFGSGMNRFRNSRTYSKFNLSLVVSFLS
 YDYYRPRFLEENRNFVSRSMVLKLTCLVRMGYQC
 TFGVLQAGQYGVAGTTRRAILAAAPGKLPFPELHVFA
 PRACOLSVVDDKFSNITRLSGSPERTTMRDTMSDLPE
 IONGASAPESLYKWRATVLPVPEAAARVALPAHPQGPYPQVH
 ERAGGCRM-->VC (in isoform 5).
 /FTid=VSP_005621.
 FT VARSPLIC 1216 1504 Missing (in isoform 6).
 /FTid=VSP_005622.
 FT VARSPLIC 1226 1477 Missing (in isoform 8).
 /FTid=VSP_005623.
 FT VARSPLIC 1252 1482 Missing (in isoform 4).
 /FTid=VSP_005624.
 FT VARSPLIC 1259 1481 Missing (in isoform 3).
 /FTid=VSP_005625.
 FT VARSPLIC 1323 1403 Missing (in isoform 7).
 /FTid=VSP_005626.
 FT CONFLICT 17 25 /FTid=VSP_005627.
 FT CONFLICT 189 189 AGSLPDHVR-->RQAPRPPCP (IN REF. 3).
 FT CONFLICT 1276 1276 A -> V (IN REF. 3).
 FT CONFLICT 1300 1300 F -> S (IN REF. 1).
 FT CONFLICT 1372 1372 T -> I (IN REF. 4).
 FT CONFLICT 1394 1428 M -> V (IN REF. 1).
 FT CONFLICT 1394 1428 KWRATVLPVPEAAARVALPAHPQGPYPQVHERAGGC -> NG
 EPOSWFQRLRGSHYQPILRDHICKMSALVAA (IN
 REF. 1).
 SQ SEQUENCE 1622 AA; 182773 MW; FCF4AAA69E234BA CRC64;
 Query Match 42.6%; Score 40.5; DB 1; Length 1622;
 Best Local Similarity 52.9%; Pred. No. 1.2e+02;
 Matches 9; Conservative 2; Mismatches 5; Indels 1; Gaps 1;
 QY 4 TANATVYMI-DSVLMPP 19
 DB 931 TKNGVYRLGDSVYLP 947
 RESULT 13
 ID Y119 YEAST STANDARD; PRT; 205 AA.
 AC P40500.
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF26772; AAG01802.1; -.
DR HAMAP; MF_00331; -; 1.
DR InterPro; IPR000192; Aminotrans V.
DR Pfam; PF00266; aminotran 5; 1._
DR PROSITE; PS00595; AA_TRANSFER_CLASS_5; FALSE_NEG.
DR Lyase; Pyridoxal phosphate.
KW ACT_SITE 338
FT ACT_SITE 338
FT SEQUENCE 404 AA; 44335 MW; 87B5BA25F87E2A25 CRC64;
SQ
Query Match 42.1%; Score 40; DB 1; Length 404;
Best Local Similarity 52.9%; Pred. No. 33;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 VSTANATVMIDSVLMP 18
||| ||| ||| ||| |||
Db 11 VSTENKAVTMDNSATTP 27

RESULT 15
ST34_ARATH
ID ST34_ARATH STANDARD; PRT; 653 AA.
AC Q9LW86;
DT 10-OCT-2003 (Rel. 42, Created)
DI 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable sulfate transporter 3.4.
GN SULTR3;4 OR AT3G15990 OR MSL1.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RP Takahashi H., Watanabe-Takahashi A., Saito K., Yamaya T.;
RL "cDNA for sulfate transporter Sultr3;4.";
RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RP MEDLINE=20277480; PubMed=10819329;
RX Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).
CC -!- FUNCTION: H(+)/sulfate cotransporter that may play a role in the
CC regulation of sulfate assimilation (by similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the SLC26A/SulP transporter (TC 2.A.53)
CC family.
CC -!- SIMILARITY: Contains 1 STAS domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB054645; BAB21264.1; -.
DR EMBL; AB012247; BAB02665.1; -.
DR InterPro; IPR002645; STAS.
DR InterPro; IPR001902; Sulph. transpt.

DR Pfam; PF01740; STAS; 1.
 DR Pfam; PF00916; Sulfate_transp; 1.
 DR TIGRFAMs; TIGR00815; sulP; 1.
 DR PROSITE; PS01130; SLC28A; FALSE_NEG.
 DR PROSITE; PS08001; STAS; 1.
 KW Transport; Symport; Sulfate transport; Transmembrane;
 KW Multigene family.
 FT DOMAIN 1 92 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 93 113 POTENTIAL.
 FT DOMAIN 114 115 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 116 136 POTENTIAL.
 FT DOMAIN 137 140 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 141 161 POTENTIAL.
 FT DOMAIN 162 167 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 168 188 POTENTIAL.
 FT DOMAIN 189 194 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 195 215 POTENTIAL.
 FT DOMAIN 216 247 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 248 268 POTENTIAL.
 FT DOMAIN 269 279 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 280 300 POTENTIAL.
 FT DOMAIN 301 331 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 332 352 POTENTIAL.
 FT DOMAIN 353 370 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 371 391 POTENTIAL.
 FT DOMAIN 392 407 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 408 428 POTENTIAL.
 FT DOMAIN 429 433 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 434 454 POTENTIAL.
 FT DOMAIN 455 471 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 472 492 POTENTIAL.
 FT DOMAIN 493 653 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 520 643 STAS.
 SQ SEQUENCE 653 AA; 71176 MW; 88FBF572E32E08BF CRC64;

Query Match 42.1%; Score 40; DB 1; Length 653;
 Best Local Similarity 56.2%; Pred. No. 55;

Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 TANATVYMIDSVLMPP 19

Db 18 TAGETVVEIHSVCLPP 33

Search completed: March 10, 2004, 12:06:36
 Job time : 6.84615 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 11:57:36 ; Search time 30.0833 Seconds
(without alignments)
199.275 Million cell updates/sec

Title: US-10-044-703-80

Perfect score: 95

Sequence: 1 GVSTANATYVIMDSVLMPP 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertibrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	95	100.0	226	Q9RMU8	Q9rmu8 mycobacteri
2	87	91.6	106	Q9RMU9	Q9rmu9 mycobacteri
3	65	68.4	219	Q9RD45	Q9rd45 streptomyc
4	65	68.4	220	Q86667	Q86667 streptomyc
5	57	60.0	623	Q9RXB6	Q9rxb6 deinococcus
6	54	56.8	803	Q8IUH0	Q8iuh0 homo sapien
7	54	56.8	2570	Q8YU15	Q8yu15 homo sapien
8	54	56.8	2570	Q8YU11	Q8yu11 homo sapien
9	54	56.8	2589	Q9J072	Q9j072 homo sapien
10	53	55.8	344	Q9GYW9	Q9gyw9 lytechinus
11	52	54.7	160	Q9R9N9	Q9r9n9 rhizobium m
12	52	54.7	160	Q922A8	Q922a8 rhizobium m
13	52	54.7	178	Q9AB20	Q9ab20 caulobacter
14	51	53.7	164	Q7U0B1	Q7u0b1 rhodospirell
15	51	53.7	206	Q83WS1	Q83ws1 microcystis
16	51	53.7	252	Q52931	Q52931 uncultured

17	51	53.7	252	2	Q52932	Q52932 uncultured
18	51	53.7	309	2	Q83V13	Q83v13 microcystis
19	51	53.7	311	2	Q83V09	Q83v09 microcystis
20	51	53.7	311	2	Q83U7	Q83u7 microcystis
21	51	53.7	312	2	Q83V08	Q83v08 microcystis
22	51	53.7	393	2	Q48915	Q48915 microcystis
23	51	53.7	1290	2	Q9SLA7	Q9sla7 microcystis
24	51	53.7	2114	2	Q8RTG4	Q8rtg4 microcystis
25	50.5	53.2	411	17	Q973M4	Q973m4 sulfolobus
26	50	52.6	252	2	Q52930	Q52930 uncultured
27	50	52.6	252	2	Q54565	Q54565 uncultured
28	50	52.6	252	2	Q52929	Q52929 uncultured
29	50	52.6	309	2	Q83V07	Q83v07 microcystis
30	50	52.6	440	11	Q8BU58	Q8bu58 mus musculu
31	50	52.6	877	17	Q8TQ29	Q8tq29 methanosarc
32	50	52.6	2571	11	Q8R4Y4	Q8r4y4 mus musculu
33	49	51.6	581	16	Q8A055	Q8a055 bacteroides
34	48	50.5	310	2	Q83V11	Q83v11 microcystis
35	48	50.5	952	2	Q845V7	Q845v7 microcystis
36	47.5	50.0	406	17	Q97UC6	Q97uc6 sulfolobus
37	47	49.5	184	16	Q8X15	Q8x15 bradyrhizob
38	47	49.5	2126	2	Q93LR1	Q93lr1 microcystis
39	46	48.4	167	16	Q877K8	Q877k8 bradyrhizob
40	46	48.4	220	16	Q8YIN4	Q8yin4 anabaena sp
41	46	48.4	226	2	Q83WM8	Q83wm8 campylobact
42	46	48.4	263	17	Q8TPB7	Q8tpb7 methanosarc
43	46	48.4	429	16	Q9PPA2	Q9ppa2 campylobact
44	46	48.4	465	10	Q9ASN3	Q9asn3 oryza sativ
45	46	48.4	683	11	P82198	P82198 mus musculu

ALIGNMENTS

RESULT 1

Q9RMU8 ID Q9RMU8 PRELIMINARY; PRT; 226 AA.
AC Q9RMU8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mpt70.
GN MPT70.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Juarez M.D., Torres A., Bigi F., Eptitia C.;
RT "Mycobacterium tuberculosis mpt83 and dipz/thioredoxin genes are part
of the same translational unit."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF189006; AAF13402.1;
DR GO; GO:0007155; P, cell adhesion; IEA.
DR InterPro; IPR000782; BIGH3_FAS1.
DR Pfam; PF02469; Fasciclin; I.
DR SMART; SM00554; FAS1; 1.
DR PROSITE; P85023; FAS1; 1.
SQ SEQUENCE 226 AA; 22467 MW; 7AB8517F63A12751 CRC64;

Query Match 100.0%; Score 95; DB 2; Length 226;
Best Local Similarity 100.0%; Pred. No. 4.4e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVSTANATYVIMDSVLMPP 19

Db 207 GVSTANATYVIMDSVLMPP 225

RESULT 2

Q9RMU9

ID Q9RM9 PRELIMINARY; PRT; 106 AA.
 AC Q9RM9;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE M83 (Fragment).
 GN M83.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RA Juarez M.D., Torres A., Bigi F., Bepitita C.;
 RT "Mycobacterium tuberculosis mpt83 and dip2/thioredoxin genes are part
 RT of the same translational unit."
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF189006; AAF13400.1;
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR000782; BIGH3 FAS1.
 DR Pfam; PF02469; Fasciclin; 1.
 DR SMART; SM00554; FAS1; 1.
 DR PROSITE; PS0213; FAS1; 1.
 FT NON TER
 SQ SEQUENCE 106 AA; 11055 MW; BE03529F3BE0CA3D CRC64;

Query Match 91.6%; Score 87; DB 2; Length 106;
 Best Local Similarity 89.8%; Pred. No. 4.7e-07;
 Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVSTANATVYMDSVLMP 19
 DB 86 GVHTANATVYMDTVLMP 104

RESULT 3
 Q9RD45 PRELIMINARY; PRT; 219 AA.
 ID Q9RD45
 AC Q9RD45;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative lipoprotein.
 GN SC0638 OR SCF56.22C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Murphy L., Harris D.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RC MEDLINE=97000351; PubMed=8843436;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RC MEDLINE=97000351; PubMed=8843436;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RC MEDLINE=97000351; PubMed=8843436;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]

RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)."
 RL Nature 417:141-147(2002).
 DR EMBL; AL939106; CAB62767.1;
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR000782; BIGH3 FAS1.
 DR Pfam; PF02469; Fasciclin; 1.
 DR SMART; SM00554; FAS1; 1.
 DR PROSITE; PS0213; PROKAR_LIPOPROTEIN; 1.
 DR Lipoprotein; Complete proteome.
 KW SEQUENCE 219 AA; 22255 MW; 863P97B1D0E4AF91 CRC64;

Query Match 68.4%; Score 65; DB 16; Length 219;
 Best Local Similarity 76.5%; Pred. No. 0.0064;
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 VSTANATVYMDSVLMP 18
 DB 201 VRTANATVYMDTVLMP 217

RESULT 4
 O86667 PRELIMINARY; PRT; 220 AA.
 ID O86667
 AC O86667;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative secreted protein.
 GN SC06375 OR SC4A2.11C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RC MEDLINE=97000351; PubMed=8843436;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,

RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RL coelicolor A3(2).";
 RT Nature 417:141-147(2002).
 DR EMBL: AL939127, CAA20163.1; -;
 DR PIR: T34986, T34986.
 DR GO: GO:0007155; P:cell adhesion; IEA.
 DR InterPro: IPR000782; BIGH3_FAS1.
 DR Pfam: PF02469; Fasciclin; 1.
 DR SMART: SMO0554; FAS1; 1.
 DR PROSITE: PS05213; FAS1; 1.
 KW Complete proteome.
 SQ SEQUENCE 220 AA; 22362 MW; 78C08D7932B12C32 CRC64;
 Query Match 68.4%; Score 65; DB 16; Length 220;
 Best Local Similarity 61.1%; Pred. No. 0.0064;
 Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 QY 2 VSTANATVYMDSVLMPP 19
 Db 191 IPTKNATVIVDGLVFP 208
 RESULT 5
 Q9XB6 PRELIMINARY; PRT; 623 AA.
 ID Q9XB6
 AC Q9XB6
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Osteoblast specific factor 2-related protein.
 DR DR0399.
 OS Deinococcus radiodurans.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 RT radiodurans R1."; Science 286:1571-1577(1999).
 RL EMBL: AB001900; AAF09979.1; -;
 DR PIR: F75523; F75523.
 DR TIGR: DR0399; -;
 DR GO: GO:0007155; P:cell adhesion; IEA.
 DR InterPro: IPR000782; BIGH3_FAS1.
 DR Pfam: PF02469; Fasciclin; 3.
 DR SMART: SMO0554; FAS1; 3.
 DR PROSITE: PS05213; FAS1; 3.
 KW Complete proteome.
 SQ SEQUENCE 623 AA; 62258 MW; AA987FF6DBB8B500 CRC64;

Query Match 60.0%; Score 57; DB 16; Length 623;
 Best Local Similarity 50.0%; Pred. No. 0.47; Length 623;
 Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 VSTANATVYMDSVLMPP 19
 Db 603 ITAGTSTVVIDTVLLPP 620

RESULT 6

QSIUHO
 ID QSIUHO PRELIMINARY; PRT; 803 AA.
 AC QSIUHO;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Soluble form FELE-1.
 GN FELE-1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22206516; PubMed=12077138;
 RA Adachi H., Tsujimoto M.;
 RT "FEEL-1, a novel scavenger receptor with in vitro bacteria-binding and
 RT angiogenesis-modulating activities."; J. Biol. Chem. 277:34264-34270(2002).
 RL J. Biol. Chem. 277:34264-34270(2002).
 DR EMBL: AB052957; BAC15607.1; -;
 DR GO: GO:0005198; P:structural molecule activity; IEA.
 DR GO: GO:0007155; P:cell adhesion; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.
 DR InterPro: IPR000782; BIGH3_FAS1.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR InterPro: IPR06209; EGF like.
 DR InterPro: IPR06210; EGF.
 DR InterPro: IPR002049; Laminin_EGF.
 DR Pfam: PF00008; EGF; 1.
 DR Pfam: PF02469; Fasciclin; 1.
 DR PRINTS: PRO0011; EGF_LAMININ.
 DR SMART: SMO0181; EGF; 6.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 DR PROSITE: PS00022; EGF_1; 2.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS02113; FAS1; 2.
 SQ SEQUENCE 803 AA; 87025 MW; 1A833922D0F223FB CRC64;
 Query Match 56.8%; Score 54; DB 4; Length 803;
 Best Local Similarity 44.4%; Pred. No. 2;
 Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 QY 2 VSTANATVYMDSVLMPP 19
 Db 627 VMAANGVIHMDGLVLP 644
 RESULT 7
 Q9NY15 PRELIMINARY; PRT; 2570 AA.
 ID Q9NY15
 AC Q9NY15;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Stabilin-1.
 GN STAB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Politz O., Guillot P., Gratchev A., Schledzewski K., Birk R.,
 RA Hakiy N., Tebbe B., Orfanos C.E., Goerdts S.;
 RT "Stabilin-1: an endothelial-macrophage member of the fasciclin domain
 RT containing protein family associated with angiogenesis."; Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ275213; CAB61827.1; -;
 DR HSP; P98066; ITSG.
 DR GO: GO:0005540; P:hyaluronic acid binding; IEA.
 DR GO: GO:0005198; P:structural molecule activity; IEA.
 DR GO: GO:0007155; P:cell adhesion; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.

DR InterPro: IPR000782; BIGH3 FAS1.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR InterPro: IPR006209; EGF_like.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR000538; Link.
 DR Pfam: PF00008; EGF; 11.
 DR Pfam: PF02469; Fasciclin; 4.
 DR Pfam: PF00193; Xlink; 1.
 DR ProDom: PD000918; Link; 1.
 DR SMART: SM00180; EGF_Lam; 1.
 DR SMART: SM00554; FAS1; 7.
 DR SMART: SM00445; Link; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; 2.
 DR PROSITE: PS00022; EGF_1; 7.
 DR PROSITE: PS01186; EGF_2; 16.
 DR PROSITE: PS0213; FAS1; 7.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 2.
 DR PROSITE: PS01241; Link; 1.
 KW EGF-like domain; Laminin EGF-like domain.
 SQ SEQUENCE 2570 AA; 275345 MW; 3123FABD7C8E2BF8 CRC64;

Query Match 56.8%; Score 54; DB 4; Length 2570;
 Best Local Similarity 44.4%; Pred. No. 7.1;
 Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSTANATVYMDSVLMPP 19
 DB 627 VMAANGVIHMDGILLPP 644

RESULT 8

ID Q8IUH1 PRELIMINARY; PRT; 2570 AA.
 AC Q8IUH1;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE FELE-1.
 GN FELE-1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2206516; PubMed=12077138;
 RA Adachi H., Tsujimoto M.,
 RT "FELE-1, a novel scavenger receptor with in vitro bacteria-binding and
 angiogenesis-modulating activities."
 RL J. Biol. Chem. 277:34264-34270(2002).
 DR EMBL; AB052956; BAC15606.1; -.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0005540; F:hyaluronic acid binding; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro: IPR000782; BIGH3 FAS1.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR InterPro: IPR001881; EGF Ca.
 DR InterPro: IPR006209; EGF_like.
 DR InterPro: IPR006210; IEGF.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR000538; Link.
 DR Pfam: PF00008; EGF; 13.
 DR Pfam: PF02469; Fasciclin; 4.
 DR Pfam: PF00193; Xlink; 1.
 DR ProDom: PD000918; Link; 1.
 DR SMART: SM00181; EGF; 23.
 DR SMART: SM00179; EGF CA; 6.
 DR SMART: SM00180; EGF_Lam; 4.
 DR SMART: SM00445; Link; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; 2.
 DR PROSITE: PS00022; EGF_1; 7.

DR PROSITE: PS01186; EGF_2; 16.
 DR PROSITE: PS0213; FAS1; 7.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 2.
 DR PROSITE: PS01241; Link; 1.
 SQ SEQUENCE 2570 AA; 275447 MW; BEF682DFC35F5CA9 CRC64;

Query Match 56.8%; Score 54; DB 4; Length 2570;
 Best Local Similarity 44.4%; Pred. No. 7.1;
 Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSTANATVYMDSVLMPP 19
 DB 627 VMAANGVIHMDGILLPP 644

RESULT 9

ID Q93072 PRELIMINARY; PRT; 2589 AA.
 AC Q93072;
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein KIAA0246 (Fragment).
 GN KIAA0246.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97191544; PubMed=9039502;
 RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
 RA Tanaka A., Kotani H., Miyajima N., Nomura N.,
 RT "Prediction of the coding sequences of unidentified human genes. VI.
 RT analysis of cDNA clones from cell line KG-1 and brain."
 RL DNA Res. 3:321-329(1996).
 DR EMBL; D87433; BAA13377.2; -.
 DR HSSP; P98066; ITSG.
 DR Genew; HGNC:18628; STAB1.
 DR GO; GO:0005540; F:hyaluronic acid binding; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro: IPR000782; BIGH3 FAS1.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR InterPro: IPR006209; EGF_like.
 DR InterPro: IPR006210; IEGF.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR000538; Link.
 DR Pfam: PF00008; EGF; 12.
 DR Pfam: PF02469; Fasciclin; 4.
 DR Pfam: PF00193; Xlink; 1.
 DR ProDom: PD000918; Link; 1.
 DR SMART: SM00181; EGF; 25.
 DR SMART: SM00445; Link; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; 2.
 DR PROSITE: PS00022; EGF_1; 7.
 DR PROSITE: PS01186; EGF_2; 16.
 DR PROSITE: PS0213; FAS1; 7.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 2.
 DR PROSITE: PS01241; Link; 1.
 KW Hypothetical protein; EGF-like domain.
 FT NON TER 1
 SQ SEQUENCE 2589 AA; 277512 MW; 86F996423001C756 CRC64;

Query Match 56.8%; Score 54; DB 4; Length 2589;
 Best Local Similarity 44.4%; Pred. No. 7.1;
 Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSTANATVYMDSVLMPP 19
 DB 627 VMAANGVIHMDGILLPP 644

Db 646 VMAANGVIHMLDGLILPP 663

RESULT 10
Q9GYW9 PRELIMINARY; PRT; 344 AA.
ID Q9GYW9
AC Q9GYW9
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 30 kDa yolk granule protein YP30.
OS Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinozoa; Echinodermata; Echinacea; Temnopneuroidea; Toxopneustidae;
OC Lytechinus.
OC Lytechinus.
OX NCBI_TaxID=7654;
RN [1]
RP SEQUENCE FROM N.A.
RA Wessel G.M., Zaydudim V., Hau J., Laidlaw M., Brooks J.M.;
RT "Direct molecular interaction of a conserved yolk granule protein in
sea urchins."
RL Dev. Growth Differ. 42:0-0(2000).
DR EMBL; AF289086; AAG02421.1; -;
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR000782; BIGH3 FAS1.
DR Pfam; PF02469; Fasciclin; 2.
DR SMART; SM00554; FAS1; 2.
DR PROSITE; PS00213; FAS1; 2.
SQ SEQUENCE 344 AA; 38621 MW; B1769D73B30BFE86 CRC64;

Query Match 55.8%; Score 53; DB 5; Length 344;
Best Local Similarity 44.4%; Pred. No. 1.2;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSTANATVYMIDSVLMPP 19
Db 320 IPTNGVHVHIDQVLLPP 337

RESULT 11
Q9R9N9 PRELIMINARY; PRT; 160 AA.
ID Q9R9N9
AC Q9R9N9
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SM12C4.2
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=1021;
RX MEDLINE=20551116; PubMed=11097914;
RA Davey M.E., de Bruijn F.J.;
RT "A Homologue of the Tryptophan-Rich Sensory Protein TspO and FixL
RT Regulate a Novel Nutrient Deprivation-Induced Sinorhizobium meliloti
RT Locus."
RL Appl. Environ. Microbiol. 66:5353-5359(2000).
DR EMBL; AF179401; AAF01193.1; -;
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR000782; BIGH3 FAS1.
DR Pfam; PF02469; Fasciclin; 1.
DR SMART; SM00554; FAS1; 1.
DR PROSITE; PS00213; FAS1; 1.
KW Hypothetical protein.
SQ SEQUENCE 160 AA; 16506 MW; 3F698A8DFF1304B2 CRC64;

Query Match 54.7%; Score 52; DB 2; Length 160;
Best Local Similarity 38.9%; Pred. No. 0.8;
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSTANATVYMIDSVLMPP 19
Db 141 IAAASNGVHVHIDKVMPP 158

RESULT 12
Q92ZA8 PRELIMINARY; PRT; 160 AA.
ID Q92ZA8
AC Q92ZA8
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nex18 symbiotically induced conserved protein.
GN NEX18 OR RA0582 OR SMA1077.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid."
RT Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
EL EMBL; AE007248; AAK65240.1; -;
DR PIR; F95334; F95334.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR000782; BIGH3 FAS1.
DR Pfam; PF02469; Fasciclin; 1.
DR SMART; SM00554; FAS1; 1.
DR PROSITE; PS00213; FAS1; 1.
KW Plasmid; Complete proteome.
SQ SEQUENCE 160 AA; 16482 MW; ASF107A3DFACAEED CRC64;

Query Match 54.7%; Score 52; DB 16; Length 160;
Best Local Similarity 38.9%; Pred. No. 0.8;
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSTANATVYMIDSVLMPP 19
Db 141 IAAASNGVHVHIDKVMPP 158

RESULT 13
Q9AB20 PRELIMINARY; PRT; 178 AA.
ID Q9AB20
AC Q9AB20
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein CC0414.
GN CC0414.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Chit N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Pladke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,

```

[1]
RN RP SEQUENCE FROM N.A.
RA Kurmayer R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
(2)
RN RP SEQUENCE FROM N.A.
RX MEDLINE=2208976; PubMed=11984633;
RA Kurmayer R., Dittmann E., Fastner J., Chorus I.;
RT "Diversity of microcystin genes within a population of the toxic cyanobacterium Microcystis spp. in Lake Wannsee (Berlin, Germany).";
RL Microb. Ecol. 43:107-118(2002).
DR EMBL; AJ539092; CAD62170.1; -;
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR00873; AMP-bind.
DR PROSITE; PS00455; AMP_BINDING; 1.
FT NON TER
FT I
FT NON TER
SQ SEQUENCE 206 AA; 22971 MW; 17A350AE918699D8 CRC64;
Query Match 53.7%; Score 51; DB 2; Length 206;
Best Local Similarity 55.6%; Pred. No. 1.6;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps
QY 1 GVSTANATVYMIDSVLMP 18
DB 169 GTPIANAQVILDSYLQP 186
Search completed: March 10, 2004, 12:11:02
Job time : 31.2083 secs

```